On the Contribution of Payments and Clearing to Financial Inclusion and Stability

Three Essays

Inauguraldissertation zur Erlangung der Würde eines

Doctor rerum oeconomicarum

der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern

vorgelegt von

Robert Oleschak
von Kreuzlingen, Thurgau

Bern, Juli 2020

Original document saved on the web server of the University Library of Bern

This work is licensed under a

Creative Commons Attribution-Non-Commercial-No derivative works 2.5 Switzerland licence.
To see the licence go to http://creativecommons.org/licenses/by-nc-nd/2.5/ch/ or write to
Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

source: https://doi.org/10.48549/2529 | downloaded: 5.5.2023

Copyright Notice

This document is licensed under the Creative Commons Attribution-Non-Commercial-No derivative works 2.5 Switzerland. http://creativecommons.org/licenses/by-nc-nd/2.5/ch/

You are free:

to copy, distribute, display, and perform the work

Under the following conditions:

Attribution. You must give the original author credit.

Non-Commercial. You may not use this work for commercial purposes.

No derivative works. You may not alter, transform, or build upon this work..

For any reuse or distribution, you must take clear to others the license terms of this work.

Any of these conditions can be waived if you get permission from the copyright holder.

Nothing in this license impairs or restricts the author’s moral rights according to Swiss law.

The detailed license agreement can be found at:
http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de

The faculty accepted this work as dissertation on 20 August 2020 at the request of the two advisors Prof. Dr. Cyril Monnet and Prof. Dr. Thorsten V. Koepp, without wishing to take a position on the view presented therein.

Acknowledgments

It was only thanks to the support of individuals and institutions that this dissertation could be accomplished. I would like to express my very great appreciation to my doctoral supervisors Prof. Dr. Cyril Monnet and Prof. Dr. Thorsten V. Koeppl for their trust, advise, and support. I am grateful for valuable discussions with my fellow doctoral candidates Fabienne Schneider, Christian Wipf, and Lukas Voellmy.

This thesis would not have been possible without the support from my employer, the Swiss National Bank (SNB). I would like to extend my thank to my colleagues Beatrice Baier, Urs Bischof, Frédéric Bos, Marco Cecchini, Andreas Fuster, Jürg Mägerle, Dave Maurer, Sarah Muster, Thomas Nellen, Felix Schafroth, Andy Sturm, the research coordination staff at the SNB, Henry Holden from the Bank for International Settlements (BIS), and seminar participants at the SNB and BIS for their insightful comments. I would also wish to thank the research staff at the Study Center Gerzensee and special thanks to Teodora Ruiz and Nina Weibel for their hospitality.

Finally, my gratitude goes to the love of my live, my wife, Rekha Oleschak, for her unfailing support, love, and understanding, for believing in me during both good and difficult phases of the project, and for providing inspiration to my research.

Bern, July 2020 Robert Oleschak

To Nihar Adrian, Emil Nishad, and Tara Nandini

Abstract

This thesis models frictions that occur in payments and clearing activities and analyzes how payments and clearing arrangements, technologies, and regulatory policies reduce or eliminate such frictions and thus contribute to financial inclusion and financial stability.

The first paper is a theoretical and empirical investigation into the trade-offs between financial inclusion, tax revenues, and the rate of inflation and how this trade-off changes with technologies providing cheaper access to payment services. The second paper theoretically analyses the incentives of a central counterparty and their surviving clearing agents during an auction of the positions of a defaulted clearing agent. The third paper studies how bilateral and central clearing arrangements deal with the moral hazard problem in markets where agents are inclined to promise excessively high contingent future payments in relation to their capital and then subsequently default.

Contents

Chapter 1

Introduction

Payments and clearing arrangements support a wide range of economic activity. Payments, the transfer of a monetary claim e.g. in the form of cash or deposit balances, typically forms one side of an exchange between two parties. Clearing, the process of transmitting, reconciling, netting, and valuing of transactions, is used in payment systems or derivatives markets to manage counterparty risks. Common to both activities is their role in helping to reduce or overcome frictions of everyday economic activities and thereby play a crucial role in a wide range of both microeconomic and macroeconomic policy issues.

Against this background, I model in three essays how payments and clearing arrangements, related technologies, and regulatory policies address frictions such as information asymmetry or access costs and thus contribute to financial inclusion and financial stability. I analyze these issues using theoretical models based on microeconomic theory and in one instance, where data is available, I test the model predictions using panel data.

In the first paper, *Financial Inclusion, Technology and the Impact on Monetary and Fiscal Policy: Theory and Evidence*, the frictions considered are costly access to payment services and the inability of governments to tax transactions paid in cash. Such frictions lead to low levels of financial inclusion, low government tax revenues and high inflation. I use a standard monetary model with banks and introduce technologies that allow for more affordable access to payment services (e.g. FinTech or mobile phone companies) and better monitoring of cash transactions for tax purposes and analyze the effect on financial inclusion, tax revenues and the welfare maximizing rate of inflation. I show theoretically that economies with inefficient technologies not only exhibit low levels of financial inclusion and tax revenues but that it is welfare improving to use inflation tax as an additional source of tax revenue. Improvements in technology lead to higher levels of financial inclusion, more tax revenues and lower optimal rate of inflation. I test this prediction using panel data for a broad set of countries. The data show a strong and robust negative link between financial inclusion and inflation and a positive link between financial inclusion and tax revenues.

In the second paper, *Central Counterparty Auctions and Loss Allocation*, the main friction is that agents bear different private costs of holding risky, non-tradeable assets. This forms the basis for trading swaps, which are subsequently cleared by central counterparties and sets the stage for auctions that central counterparties conduct when an agent defaults. Central counterparties typically have sweeping powers in the form of recovery measures to allocate losses to surviving agents. I discuss the optimal loss allocation and show that central counterparties can increase the expected revenue of the auction but that the size of the loss for the financial system as a whole is not affected by the loss allocation. I show that while recovery measures do increase the safety and soundness of central counterparties, they may incentivize the central counterparty to delay of the auction, increasing the private costs for surviving agents.

Finally, in the third paper, *Markets in the Presence of Moral Hazard, Limited Liability and Nonexclusive Contracts*, the friction is that agents might find it profitable to trade too many contracts, exert low levels of effort and subsequently default. I show that in case of bilateral clearing, collateral solves this moral hazard problem. However, due to positive externalities, agents often prefer not to collateralize contracts. I discuss the contribution of contractual innovations (e.g. termination clauses) in solving the moral hazard problem and show their limitations. Alternatively, central counterparty clearing deals with this moral hazard problem by setting positions limits and offering default-free contracts for low levels of collateral. However, this does not work when several central counterparties clear the same market. I argue that regulatory rules on minimum level of collateral are the most robust and effective way of dealing with moral hazard issues.

Chapter 2

Financial Inclusion, Technology and the Impact on Monetary and Fiscal Policy: Theory and Evidence

Financial Inclusion, Technology and their Impacts on Monetary and Fiscal Policy: Theory and Evidence*

Robert Oleschak
Swiss National Bank (SNB)
Bundesplatz 1, CH-3003 Bern
robert.oleschak@snb.ch

June 19, 2020

Abstract

In economies with a low level of financial inclusion (FI), most activities are settled in cash and are thus more difficult to trace, record, and tax. I show theoretically that economies with inefficient financial technologies exhibit low levels of FI and of tax revenue and that using an inflation tax as an additional source of income improves welfare. Improvements in technology lead to a higher level of FI, increased tax revenue and lower (optimal) inflation. I test this prediction using panel data from a broad set of countries. The data show a strong and robust negative link between FI and inflation and a positive link between FI and tax revenue.

Keywords: Financial Inclusion, Financial Technology, Monetary Policy, Fiscal Policy

JEL Classification: C12, C22, E31, E41, G21, H21

*I would like to thank Andreas Fuster, Thorsten V. Koeppl, Cyril Monnet, Andy Sturm, Haoxiang Zhu, and an anonymous referee for their insightful comments. Disclaimer: The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the author. They do not necessarily reflect the views of the Swiss National Bank (SNB). The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

1 Introduction

The level of financial inclusion[1] has increased considerably in the last few years, supported by public and private initiatives. According to the Global Financial Inclusion Database[2], between 2011 and 2017, the global share of adults with an account at a financial institution rose from 51 to 69 percent.[3] In India, for example, the share of adults with an account more than doubled during the same time period to 80 percent. According to Demirguc-Kunt et al. (2018) an important driving factor was government-issued biometric identification cards, lowering the cost of access and boosting account ownership among unbanked adults. D'Silva et al. (2019) provide a detailed case study of how India's provision of digital financial infrastructure contributed to financial inclusion.

More generally, new applications of technology to financial services and to tax collection have accelerated in the last few years with implications for how financial services are provided, accessed and used and how effectively taxes are collected. CPMI-IOSCO (2020) provides a recent account of the opportunities and risks in fintech developments for financial inclusion.[4] The report stresses that to harness this technology, the government needs to provide financial and ICT infrastructure, legal and regulatory frameworks and collaborate with the private sector. Similar approaches are advocated to increase the effectiveness of tax collection by, for example, enhancing the ability of tax authorities to detect economic activity in the informal sector (see, for example, Bird and Zolt (2008)). Countries that successfully provide the foundations for harnessing new technologies are in a good position to improve both financial inclusion and tax collection.

The growth in account ownership has not been uniform across countries. The share of adults with a transaction account at a financial institution in a country can vary from slightly above zero to 100 percent. Economies with low levels

[1] Financial inclusion generally means that individuals and businesses have access to useful and affordable financial products and services that meet their needs—for transactions, payments, savings, credit and insurance—delivered in a responsible and sustainable way. See https://www.worldbank.org/en/topic/financialinclusion/overview

[2] https://globalfindex.worldbank.org

[3] Individuals with an account can make and receive cashless payments or store value.

[4] These range from application programming interfaces, big data analytics, cloud computing, contactless technology, digital identification for distributed ledger technologies and the Internet of Things leading to new products and services.

of financial inclusion settle most economic activities in cash, which implies that these activities are more difficult to trace, record, and tax. Accordingly, in order to finance public spending, governments often need to resort to other sources of funding. For economies that rely heavily on cash, governments use the inflation tax (or seigniorage taxation) as an additional source of income.

Figure 1, which plots the share of adults with an account at a financial institution against inflation and tax revenue (as a percent of GDP), supports this claim: the two panels show a negative correlation between financial inclusion and inflation and a positive correlation between tax income and financial inclusion.[5] What does the literature have to say about this?

Given the amount of attention financial inclusion has received in the last decade, there are surprisingly few studies that explore, either theoretically or empirically, whether governments adjust inflation and taxation to the level of financial inclusion. Several studies consider the somewhat related impact of the shadow economy on tax revenue and inflation either from a theoretical point of view (for example, Koreshkova (2006) or Nicolini (1998)) or use estimates to predict the impact of the shadow economy on taxes and inflation (Mazhar and Meon (2017)). In addition, Roubini and Sala-i-Martin (1995) show in a theoretical model that financial repression is associated with high levels of tax evasion, low growth, and high inflation and Levine (1997) discusses more generally the relationship between financial development and economic growth. To my knowledge, only Oz-Yalaman (2019) assesses the impact of financial inclusion on tax revenue and finds a significant and positive relationship between financial inclusion and tax revenue. A theoretical and empirical approach to financial inclusion and both tax revenue and inflation is still missing.

In this paper, I aim to fill this void by analyzing the relationship between financial inclusion, tax revenue, and inflation theoretically as well as empirically and setting it in a broader technological context. Building on Lahcen and Pedro (2019), the theoretical monetary model allows households to decide endogenously whether to join the financial system or not, i.e., where financial inclusion is an equilibrium

[5] Of course, the figure does not reflect many other factors that affect tax revenue and inflation, e.g., differences in institutional quality. This will be controlled for in the empirical portion of the paper.

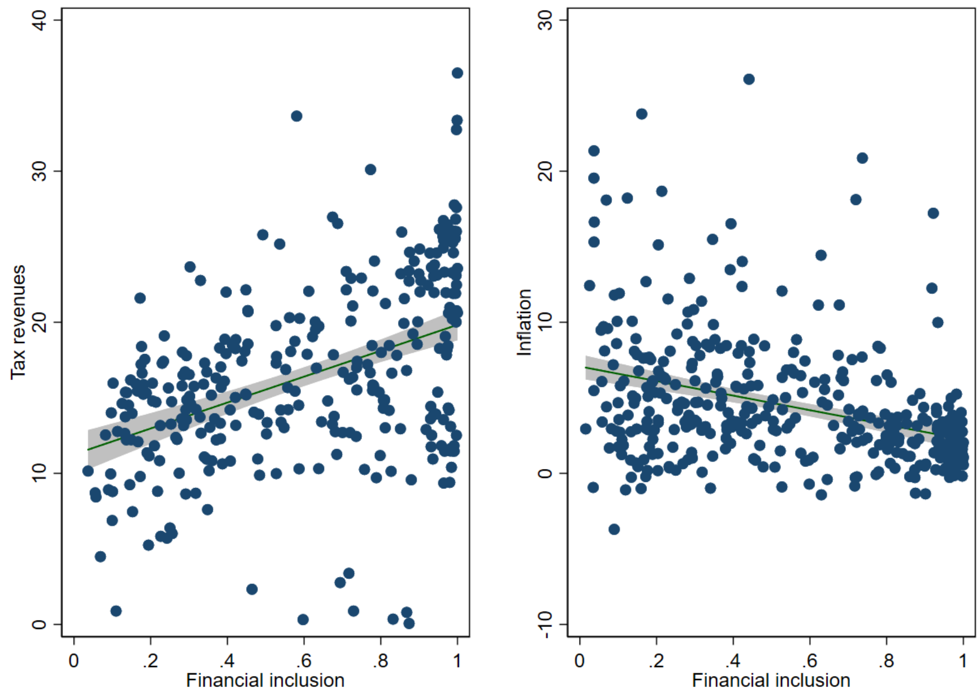


Figure 1: Financial Inclusion (share of adults with an account at a financial institution), Inflation (Consumer Price Index), and Tax Revenue (as a % of GDP), 2011, 2014, 2017

outcome. Inflation, taxes, interest rates on deposits, the cost of handling cash, the technology in the financial system (modeled as the utility cost of joining the financial system) and the technology for tax collection are important determinants of the equilibrium. I show that it is optimal for the government to set an inflation rate above zero (which is consistent with the findings of Koreshkova (2006) and Nicolini (1998)) and that more efficient technologies help to increase the level of financial inclusion, and simultaneously lead to a decrease in the optimal level of inflation and an increase in tax revenue. It is worth noting that such an outcome is not self-evident: an increase in inflation raises the cost of holding cash and should lead to an increase in financial inclusion, i.e., financial inclusion should increase with inflation. The model predicts the opposite, which sets the stage for the empirical analysis.

Estimating these effects empirically is challenging because technological progress in the financial system and in tax collection is difficult to observe and other policy variables such as inflation and taxes are set at the same time, together with financial inclusion. For the empirical analysis, I exploit the fact that recent technological innovations have swiftly improved access to financial services by offering better and cheaper services. For example, the entry of FinTech and mobile phone companies offering payment services—and increasingly saving, credit, and insurance services—has been swift in several countries and its potential to accelerate financial inclusion is well recognized (see, for example, Bech and Hancock (2020), Philippon (2020)). According to Frost et al. (2019) the success of these new payment service providers seems to rely on the use of technology to offer cheaper services, the ability to reach a wide audience through existing platforms and thereby reaping network effects, and the ability to process and analyze data to improve services and benefit customers even further.

In line with my hypothesis, I find a significant and robust negative relationship between financial inclusion and inflation and a positive relationship between financial inclusion and tax revenue, even after controlling for major macroeconomic variables. The relationship between inflation and financial inclusion holds even when controlling for the independence of central banks but disappears for developed countries. The relationship between taxation and financial inclusion is more robust and holds for all specifications.

Related Literature

This paper builds on several strands of literature. First, the monetary model is based on Lagos and Wright (2005) and Rocheteau and Wright (2005), who introduce an environment where currency is essential by facilitating the exchange of goods. Many subsequent papers have expanded this core model by introducing additional features (see Lagos et al. (2017) for an overview). Most relevant to this paper, Berentsen and Waller (2007) introduced the banking sector into the standard monetary model. This paper borrows many features from Williamson (2012), who explicitly models an environment in which agents pay in currency or with interest-earning bank deposits. This feature is crucial because agents who join banks can avoid the use of cash and the affiliated costs of inflation.

Second, this paper relates to the vast literature that studies the shadow or underground economy.[6] One of the recurring findings in this literature is that shadow economy activities are found in all economies but that their size varies considerably (see Schneider and Enste (2000) for an extensive review of this literature). Some of the papers that seek to explain these variations are based on the centuries-old notion that inflation, or more generally, debasing the value of money, can be used to finance government expenditure (see, for example, Sussman (1993) for an account of France's debasement of coinage to increase revenue during the Hundred Years' War in the 15th century). For example, Gomis-Porqueras et al. (2014) use a monetary model in the tradition of Rocheteau and Wright (2005) in which agents can choose to use cash to pay for goods and avoid taxes or to use readily available credit but be charged with taxes. Since inflation increases the cost of holding cash, it provides incentives to pay with credit and thus reduces the size of the non-taxed or shadow economy; i.e., they establish that there is a negative relation between inflation and the size of the shadow economy. Nicolini (1998) and Koreshkova (2006) take the public finance perspective and argue that it might be welfare improving to use inflation to extract tax revenue when an economy is faced with a large informal sector. Nicolini (1998) takes the size of the informal economy as given in contrast to Koreshkova (2006), where the size of the formal and informal sectors is driven by the productivity gap between the two producing sectors. The financial sector in Koreshkova (2006) offers protection from inflation but does not help the government collect taxes. Thus, a more productive banking sector increases the size of the informal sector and the level of optimal inflation. Finally, Aruoba (2018) explains cross-country differences in inflation, tax revenue and the size of the informal sector through the institutional factors of a country, modeled as the difficulty of tax evasion.[7] This paper differs from this literature in several ways. In Gomis-Porqueras et al. (2014), the financial system is fully developed and efficiently operated; thus, technological changes, unlike in the present paper,

[6] Schneider and Enste (2000) defines the shadow economy as either all legal economic activities that avoid taxation or all illegal activities that avoid regulation or laws.

[7] Similarly, Ihrig and Moe (2004) and Prado (2011) seek to explain the differences in the size of the informal sector based on taxation and tax enforcement without considering inflation as a main driving force.

do not play a role. In Koreshkova (2006), the financial system helps agents avoid the cost of inflation, but the financial system does not support the government in collecting taxes. Thus, relative improvements in the efficiency of the financial system reduce tax evasion and lead to a higher level of optimal inflation. This contrasts with the role of the financial sector in this paper, in which improvements in technology (and thus efficiency) increase financial inclusion and tax revenues but decrease (optimal) inflation. In the empirical analysis, I show that as predicted by the model, there is indeed a negative relationship between financial inclusion and inflation and a positive relationship between financial inclusion and tax revenue.
Third, I model public finance trade-offs in an environment with endogenous financial inclusion and follow the tradition of Chatterjee and Corbae (1992), Franklin Allen (1994) and others who show that endogenous decisions to join financial markets have potentially rich implications for macroeconomic behavior. Lahcen and Pedro (2019) study the effects of endogenous financial inclusion on inequality and find that in such a setting, monetary policy has distributional consequences. In this paper, I find similar rich and contrasting macroeconomic trade-offs.

2 The Model

2.1 Private Economic Agents

Following Lagos and Wright (2005) and Rocheteau and Wright (2005), at each date $t = 0, 1, ...$, agents convene sequentially in a decentralized market (DM) and a centralized market (CM). In the DM, some agents, called *sellers*, provide a good that is demanded by other agents, called *buyers*. In the CM, both buyers and sellers meet in a centralized Walrasian market and make production and consumption decisions. Let there be a continuum of sellers and buyers, each with mass one. The period utility for buyers and sellers is given by

$$\mathcal{U}(h, q) = -h + u(q) \quad \text{and} \quad \tilde{\mathcal{U}}(x, H) = x - H$$

where the pair $\langle h, x \rangle$ represents labor and consumption in the CM, and $\langle q, H \rangle$ is consumption and labor in the DM. The utility of the buyer in the DM $u(\cdot)$ is twice continuously differentiable so that some $\hat{q} > 0$ exists, such that $u(\hat{q}) - \hat{q} = 0$. Define q^* by $u'(q^*) = 1$ and define the utility function as having constant relative risk aversion, or $-x\frac{u''(x)}{u'(x)} = 1$. The production technology available to buyers and sellers allows the production of one unit of the perishable consumption goods q and x, respectively, for each unit of labor supply, which hinders the barter between agents across the CM and DM. I also assume that agents are anonymous in the DM, which hinders unsecured credit. These two frictions (perishable goods and anonymity) generate a role for assets in the facilitation of exchange.

There are two assets that can serve in this capacity: currency, with supply M and ownership claims on a financial intermediary promising to pay q_{+1} units of currency in the next period with supply A. If the financial intermediary is a bank, then the claim is also referred to as a deposit. In this paper, I use the terms bank and deposits, which include any other financial entity, including Fintech and Big Tech companies. Given that ϕ is the CM price of money in terms of the consumption good x, then the real value of currency and deposits is $m = \phi M$ and $a = \phi A$. All agents can hold and receive currency, but to hold and receive deposits, agents need to join a bank; i.e., agents cannot hold government bonds directly.

In the CM, all sellers, buyers, and the government meet in a centralized Walrasian market, where production and consumption decisions are made and where buyers decide whether to join a bank and deposit their savings at that bank. Given that a fixed exogenous fraction $\rho \in [0, 1]$ of sellers have a bank account (accepting cash as well as deposits as payment), an endogenous fraction $\gamma \in [0, 1]$ of buyers decide to join a bank. A buyer's decision about whether to join a bank is affected by the following factors: the cost of opening a bank account ω, the cost of handling cash c, inflation μ, and tax payment τ, which will be discussed in detail in section 3.

Finally, in the DM, each buyer is matched at random with a seller. In cases where both the buyer and seller have joined a bank, a communication technology is available that permits the buyer to transfer ownership of a claim on the bank to the seller. Following Williamson (2012), I refer to these as *monitored* meetings. In all other cases, *nonmonitored* meetings occur, in which only currency issued by the government can be used to pay for the exchange of goods. Finally, let us assume that, when a buyer meets a seller, the buyer makes a take-it-or-leave-it offer of assets in exchange for goods.

2.2 Technology

In addition to the factors mentioned above, technology affects γ, the share of financially included buyers, in two ways. First, ω reflects the cost of accessing and using a bank account for making payments. Assuming that the banking sector is competitive and that the cost of providing a bank account is zero, ω can be interpreted as the utility cost to the buyer of opening and accessing a bank account to make payments. Thus, ω is agent specific and permanent. I assume that this utility cost is uniformly distributed across buyers, $\omega \sim U[0, k]$, where $k \in [0, 1]$ represents the *technology* in the financial sector: the lower k is, the more efficient the financial sector is.

Second, a buyer's transactions are easier to trace, record, and tax if he or she has joined a bank. Therefore, a buyer who has joined a bank pays the full tax τ, whereas a buyer who has not joined a bank pays only $(1-\theta)\tau$, where $(1-\theta) \in [0, 1]$ represents the likelihood of being caught. Improvements in the *technology* of the taxation system help increase the likelihood of being caught, i.e., the level of enforcement. This is similar to Ihrig and Moe (2004), who model the level of enforcement as the probability of being caught and then having to pay taxes or, alternatively, having to pay an additional penalty. They note that the level of enforcement is difficult to measure and use the inverse of seigniorage as a proxy. Unlike Ihrig and Moe (2004), who assume the relationship between seigniorage and enforcement, I aim to establish this relationship by arguing that technology plays an important role.

Finally, I assume that the two types of technology are closely intertwined and move in the same direction, which means that technological improvements reduce the cost of access to the financial system and support better tax enforcement; i.e., technological improvements lower the parameters k and θ.

2.3 Government

The government is a consolidated entity, consisting of a fiscal and a monetary authority. The government can levy lump-sum taxes on financially included buyers in the CM, with τ denoting the tax per buyer in units of goods used to finance real government spending G. In addition, the government issues M units of currency and B one-period nominal bonds held by the banks. The bonds issued by the government have a payoff of q_{+1} in the next period as measured in units of money. Letting ϕ denote the price of money in terms of goods in the CM market, the consolidated government budget constraint is

$$\phi(M+B)+\gamma\tau+(1-\theta)(1-\gamma)\tau=\phi(M_{-1}+qB_{-1})+G \tag{1}$$

Equation (1) states that the government's outstanding liabilities at the end of the CM, plus tax revenue, must equal the government's net outstanding liabilities at the beginning of the CM, for all periods.

To limit the class of monetary policies, I follow Williamson (2012) and fix two parameters: first, the total stock of government liabilities grows at a constant rate μ, and the ratio of currency to total government debt is a constant δ, i.e., $M=\delta(M+B)$. Since I consider only cases where the government is a net debtor ($B>0$), it follows that $\delta\in[0,1)$.

Given this, lump sum taxes can be passively determined as

$$\tau=\frac{1}{1-\theta(1-\gamma)}\Big[-\frac{\phi M}{\delta}\frac{\mu-1}{\mu}+\frac{\phi M}{\mu}\frac{1-\delta}{\delta}(q-1)+G\Big] \tag{2}$$

In equation (2), the first term in front of the square bracket expresses the scope of buyers covered by the tax; i.e., more buyers joining the financial system γ or improvements in the tax technology θ lower the lump sum tax. The first term in the square bracket expresses the negative of the proceeds from the increase in currency, i.e., higher proceeds decrease the lump sum taxes. The second term in the square bracket is the real value of the net interest on government liabilities, which needs to be financed by the lump sum tax. Finally, the last term represents the to-be-financed real government spending.

2.4 Banks

Banks form in the CM before buyers know whether they will meet a financially included or excluded seller in the DM, and dissolve in the CM of the subsequent period, when they are replaced by new banks.

Banks can invest deposits into government-issued bonds or into currency. Since I assume that the banking system is competitive, the gross real interest rate paid on the deposits equals the gross real interest rate r_{+1} of the government bonds, which is defined to be $r_{+1} = q_{+1}\frac{\phi}{\phi_{-1}}$.

When a financially included buyer meets a financially included seller, then she can pay with her interest-bearing deposits. However, if she meets a financially excluded seller, she would need to withdraw cash and forego the interest on the deposit. Following Williamson (2012), banks offer a deposit contract that maximizes the expected utility of each of its depositors by offering interest rates on deposits and minimizing the amount of cash that must be put aside to pay for transactions with financially excluded sellers. Based on this optimal offer, buyers decide whether to join the bank (and make deposits in the form of goods).

3 Equilibrium

In this section I define and characterize the equilibrium.

3.1 Financially Excluded Buyers

The CM and DM value functions are denoted by $W(m_e)$ and $V(m_e)$, where m_e refers to the currency that will be used by the financially excluded buyers in real terms. The CM problem is

$$W(m_e) = \max_{\hat{m}_e,h} -h + \beta\hat{V}(\hat{m}_e), \quad \text{s.t. } h = \mu\hat{m}_e - m_e + c + (1-\theta)\tau$$

where $\hat{m}_e$ is the real value of money taken out of the CM and put into the DM in the next period, and c is the cost of handling cash, which represents the cost of theft protection, the inconvenience of handling banknotes and coins, etc. The first-order condition (FOC) is $\beta\hat{V}'(\hat{m}_e) = \mu$ and the envelope condition $W'(m_e) = 1$ demonstrates that $W(m_e)$ is linear.

Let q_e denote the exchanged good and p_e the respective payment in a nonmonitored meeting in the DM. Then, the value function in the DM can be written as:

$$\hat{V}(\hat{m}_e) = \hat{W}(\hat{m}_e - p_e) + u(q_e), \quad \text{s.t. } p_e \leq \hat{m}_e \tag{3}$$

Since in the baseline model, we are considering take-it-or-leave-it offers, the buyer will offer the seller enough currency to cover her costs; i.e., to get q_e, the buyer offers to pay $p_e = \hat{m}_e = q_e$.

As usual, $\frac{\beta}{\mu} < 1$ implies that buyers spend all their currency when they meet a seller in the DM, i.e., that they are constrained in a meeting and that $m_e = 0$. In the case where $\frac{\beta}{\mu} = 1$, the buyer can consume the optimal amount of goods q^* and is indifferent to carrying currency into the CM or not (i.e., $m_e \geq 0$). Differentiating (3), and using the FOC from the CM, we get the Euler equation:

$$u'(\hat{m}_e) = \frac{\mu}{\beta} \tag{4}$$

Since $q_e = \hat{m}_e$, it follows that $q_e < q^*$ when $\frac{\mu}{\beta} > 1$. The discounted utility of the financially excluded buyer in the CM expressed in terms of real goods is then

$$\mathcal{V}_e = \mathcal{L}_e - c - (1-\theta)\tau, \quad \text{where } \mathcal{L}_e = -\mu q_e + \beta u(q_e) \tag{5}$$

The first term represents the net utility of the financially excluded buyer: to buy q_e goods in the DM next period, the buyer needs to acquire $\hat{m}_e = \mu q_e$ money in real terms in the CM.

3.2 Financially Included Buyers

For each buyer, the bank acquires m_n units of currency (to be spent in a nonmonitored meeting) and a units of interest-bearing assets. Since a fraction of buyers γ join a bank, all banks together acquire γm_n units of currency and γa units of interest-bearing assets.

When buyer-depositors learn their types, at the end of the CM, each depositor who will be in a nonmonitored meeting in the DM withdraws $\frac{m_n'}{1-\rho}$ units of currency. Depositors in monitored meetings each receive the right to trade away deposit claims on $\frac{(m_n - m_n' + a - a')}{\rho}$ units of the bank's original assets.[8] The CM problem for a buyer is

$$W(a) = \max_{\hat{m}_n, \hat{a}, h} -h + \beta \hat{V}(\hat{m}_n, \hat{a}), \quad \text{s.t. } aq + h = \mu(\hat{m}_n + \hat{a}) + \tau + (1-\rho)c + \omega \quad (6)$$

It is optimal for the financially included buyer to spend all currency in the DM. Therefore, the agent will only take deposit claims a into the CM. The key FOCs are $\beta \hat{V}_1(\hat{m}_n, \hat{a}) = \beta \hat{V}_2(\hat{m}_n, \hat{a}) = \mu$ and the EC is $W'(a) = q$. The DM problem can be stated as follows:

$$\hat{V}(\hat{a}, \hat{m}_n) = (1-\rho)u^n(q_n) + \rho u^m(q_m) + \hat{W}(\hat{a}') \quad (7)$$

where $q_n = \frac{\hat{m}_n'}{1-\rho}$ and $q_m = q_{+1}\frac{\hat{a}-\hat{a}'}{\rho} + \frac{\hat{m}_n - \hat{m}_n'}{\rho}$.
Naturally, q_n represents the exchanged good in a nonmonitored meeting where currency was used, and q_m is the quantity of exchanged goods in a monitored meeting against bank-deposits. Substituting (7) into (6), and expressing m and a in current real values, we can reformulate the problem of the bank as follows:

$$\max_{m_n, m_n', a, a'} -(m_n + a) + \beta(1-\rho)u^n\left(\frac{m_n'}{\mu(1-\rho)}\right) + \beta\rho u^m\left(r_{+1}\frac{a-a'}{\rho} + \frac{m_n - m_n'}{\mu\rho}\right) + \beta r_{+1} a' \quad (8)$$

The FOCs are as follows, where $q_n = \frac{m_n'}{\mu(1-\rho)}$ and $q_m = r_{+1}\frac{a-a'}{\rho} + \frac{m_n - m_n'}{\mu\rho}$:

$$\beta u^m_{m_n} - \mu = 0 \quad (9a)$$

$$u^n_{m_n'} - u^m_{m_n'} = 0 \quad (9b)$$

$$\beta r_{+1} u^m_a - 1 = 0 \quad (9c)$$

$$u^m_{a'} - 1 = 0 \quad (9d)$$

An equilibrium in which real bonds are plentiful requires that $\beta r_{+1} = 1$, in which case, according to $(9c)$ $u^m_a = 1$ and since $u^m_a = u^m_{a'}$, according to $(9d)$, the bank is willing to acquire any amount of additional real bonds a' that are available in the market. Buying more bonds than necessary for the exchange (i.e., $a' > a$) does

[8] It is natural to assume that buyers spend all currency in the DM, since we look at equilibria where $\frac{\phi_{+1}}{\phi}\beta \leq 1$.

not affect the utility of the agent.

With regard to the currency acquired by the bank, which is based on (9a) $u^m_{m_n} < u^m_a$ as long as $\mu > \beta$, and so $m_n = m'_n$. In this case, (9b) can be simply expressed as $u^n_{m'_n} = \frac{\mu}{\beta}$. If $\mu = \beta$, the bank acquires any currency that is issued by the government without affecting the utility of the agent.

Finally, the expected utility of a financially included buyer in each period is then

$$\mathcal{V}_i(\omega) = (1-\rho)\mathcal{L}_n + \rho\mathcal{L}_m - (\omega + (1-\rho)c + \tau) \tag{10}$$

where $\mathcal{L}_n = -\mu q_n + \beta u^n(q_n)$ and $\mathcal{L}_m = -\frac{1}{r_{+1}} q_m + \beta u^m(q_m)$. The first and second parts sum up the net utility in the case of a nonmonitored and a monitored meeting, respectively. The third part contains the expected costs after joining a bank $\omega + (1-\rho)c$ and the tax payment τ.

A buyer will join the financial system if $\mathcal{V}_i(\omega) \geq \mathcal{V}_e$. The share of buyers joining the financial system can be defined as $\gamma = \frac{\tilde{\omega}}{k}$ where $\mathcal{V}_i(\tilde{\omega}) = \mathcal{V}_e$.

3.3 Characterization

I confine attention to stationary monetary equilibria where real quantities are constant over time, i.e., $m = \phi M = \phi_{-1} M_{-1}$ and $b = \phi B = \phi_{-1} B_{-1}$. This implies that $\mu = \frac{\phi_{-1}}{\phi}$ and the nominal gross interest rate can be expressed as $q = \mu r$. Further, I assume that the supply of government bonds is plentiful so that, as will be shown later, the gross real interest rate is $r = \frac{1}{\beta}$.

Definition 1 *Given monetary policy μ, δ and share ρ of sellers joining banks, an equilibrium consists of real quantities of currency $m = \gamma m_n + (1-\gamma) m_e$ and plentiful amount of real bonds $b \geq \gamma a$ (to be defined below) with a gross real interest rate $r = \frac{1}{\beta}$, such that (i) m_n and a solve (8) and m_e solve (4), (ii) the tax rate is defined (based on equation (2)) as*

$$\tau = \frac{\gamma m_n + (1-\gamma) m_e}{\gamma + (1-\theta)(1-\gamma)} \left(\frac{r - \delta r - 1}{\delta} + \frac{1}{\mu} \right) + \frac{G}{\gamma + (1-\theta)(1-\gamma)} \tag{11}$$

and (iii) the share γ of buyers joining a bank is defined as $\gamma = \frac{\tilde{\omega}}{k}$ where $\mathcal{V}_i(\tilde{\omega}) = \mathcal{V}_e$.

As discussed in Williamson (2012), there are four possible equilibria: i) liquidity trap, ii) plentiful interest-bearing assets, iii) scarce interest-bearing assets, and iv) the Friedman rule. Equilibria i) and iii) require that interest-bearing assets are scarce, equilibrium ii) requires that their supply is plentiful, and equilibrium vi) is possible in a scarce or plentiful environment. Since I define the interest-bearing assets to be plentiful, I reduce the possible solutions to equilibria ii) and iv), which are characterized and discussed in the following.

3.3.1 Plentiful Interest-Bearing Assets Case

In this equilibrium, $\frac{1}{\mu} < r = \frac{1}{\beta}$, which means that the nominal interest rate on interest bearing assets is positive and that currency is comparably scarce. Therefore, based on the first-order condition for problem (8), we have $m_n' = m_n$, $a \in [\beta\rho q^*, \infty]$, and $a' \geq 0$, and m_n solves

$$\frac{\beta}{\mu} u' \Big(\underbrace{\frac{1}{\mu} \frac{m_n}{1-\rho}}_{q_n} \Big) = 1 \tag{12}$$

which allows us to make two observations. First, in nonmonitored meetings, exchange is not efficient and leads to the same size of goods exchanges as in the case of the financially excluded buyer, i.e., $q_n = q_e = q$, which implies that $m_n = (1-\rho)m_e$. Second, the assumption that $-x\frac{u''(x)}{u'(x)} = 1$ implies that the demand for currency m_b is independent of μ.

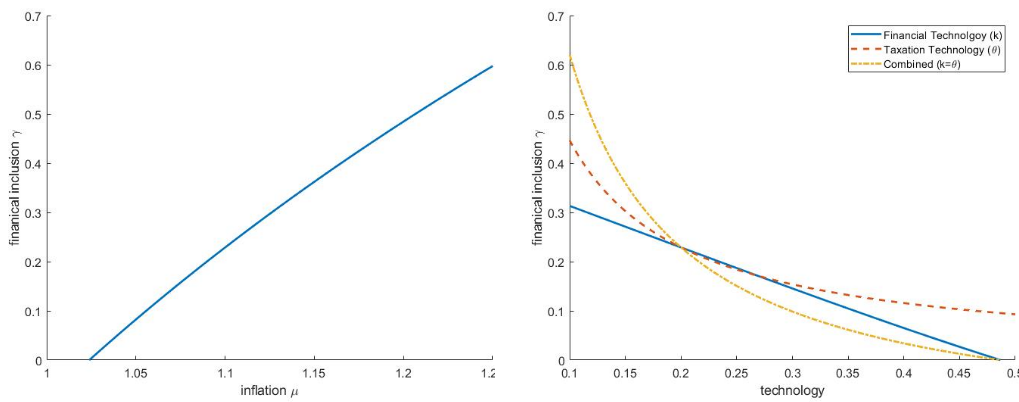
Since the rate of return on interest-bearing assets is equal to the rate of time preference, exchange is efficient in monitored meetings (all buyers receive q^* in the DM), and the bank is willing to acquire an unlimited quantity of interest-bearing assets. For the interest-bearing assets to be plentiful, we require that the supply of real bonds b is above the threshold $\underline{b} \geq \beta\gamma\rho q^*$.

The share of buyers joining a bank can be expressed as[9]

$$\gamma = \frac{\rho}{k}(\mathcal{L}^* - \mathcal{L} + c) - \frac{\theta}{k}\tau \tag{13}$$

[9]For certain parameters the implicit function, (13) has two possible solutions for $\gamma_{1,2} \in (0, 1)$. I consider only the solution leading to a higher level of financial inclusion, since the lower solution has no meaningful economic interpretation. For example, better technology would lead to a lower level of financial inclusion.

Figure 2: Drivers of Financial Inclusion



(a) Inflation and financial inclusion (b) Technology and financial inclusion

The utility $u(x) = \log(x)$ and parameters displayed in table (1) are used. In addition, for the right-hand figure, an inflation rate of $\mu = 1.1$ is applied, and for the left-hand figure, the technology parameters $k = \theta = 0.2$ are applied.

where $\mathcal{L}^* = -\beta q^* + \beta u(q^*)$, $\mathcal{L} = -\mu q + \beta u(q)$ and τ based on (11) can be rewritten as follows:

$$\tau = m_e \frac{1 - \gamma\rho}{1 - \theta(1-\gamma)} \Big(\frac{r - \delta r - 1}{\delta} + \frac{1}{\mu} \Big) + \frac{G}{\gamma + (1-\theta)(1-\gamma)} \tag{14}$$

Clearly, joining the financial system is more attractive when the share of sellers in the financial system ρ and the cost of handling cash c are high.
The effects of inflation μ and technology k as well as of θ are not straightforward because they also affect the level of taxes τ. Generally, an increase in μ boosts the level of financially included buyers γ. Similarly, a reduction in k and θ, e.g. through a cost-cutting technology allowing cheaper access to banks and better collection of taxes from the financially excluded, increases financial inclusion, as displayed in Figure (2). With growing financial inclusion, the threshold $\underline{b}$ at which interest-bearing assets are plentiful also increases.

3.3.2 Friedman Rule Case

If $\frac{1}{\mu} = r = \frac{1}{\beta}$, then $m_n > (1-\rho)q^*$, $m_n' = (1-\rho)q^*$, $a' = a + m_n - q^*$, and $a + m_b \geq q^*$.

In this case, all rates of return are equal to the rate of time preference, $\mu = \beta$, and the equilibrium exists for any number of real bonds b. Only buyers who have a cost of handling cash high enough to offset the higher tax burden join the financial system, i.e.,

$$\gamma = \frac{\rho}{k}c - \frac{\theta}{k}\tau$$

Note that the financially included would need to finance not only government spending but also the reduction in the money base. Therefore, for a certain set of parameters, there is no solution; i.e., no one joins the financial system.

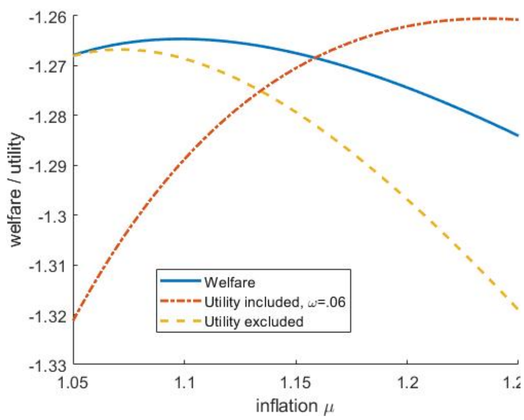
3.4 Welfare Analysis

I weight the utilities of buyers and sellers equally. In this case, the sellers drop out (zero net benefit), and welfare can be measured as the sum of the utility of financially included and excluded buyers:

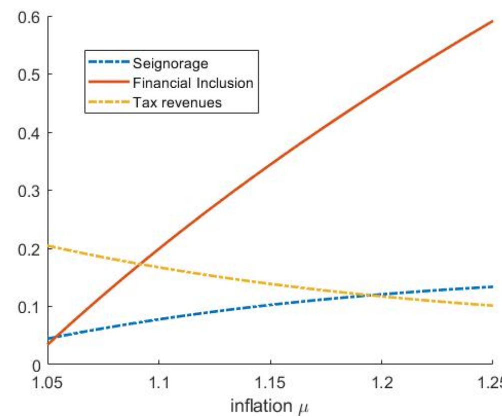
$$W = \int_0^{\omega=k\gamma} \mathcal{V}_i(x)\frac{1}{k}dx + (1-\gamma)\mathcal{V}_e$$

When comparing the plentiful interest-bearing equilibrium with the Friedman rule equilibrium, the trade-off between the level of consumption and the cost of handling cash must be considered. While the Friedman rule allows all agents to consume the same (optimal) level of goods in the DM, it is associated with a low level of financial inclusion, leading to high costs for handling cash and higher taxes for those who join the financial system. The plentiful interest-bearing equilibrium leads to a sub-optimal level of consumption among the financially excluded (and the nonmonitored meetings of the included) and thus to higher levels of inequality, but it induces higher levels of financial inclusion and thus lower costs for handling cash. This raises the question of whether welfare can be improved by introducing a wedge between the rate of inflation and real interest rates, inducing more buyers to join the banking system.

Figure 3: Welfare, seignorage and taxes



(a) Welfare and Utility



(b) Seignorage and taxes

The utility $u(x) = \log(x)$ and parameters displayed in table (1) are used. In addition, the technology parameters $k = \theta = 0.2$ are applied. Seignorage is defined as the government income gained from increasing the money base, which must equal spending net of tax income, i.e., $\phi(M - M_{-1}) = G + \phi(B - qB_{-1}) - (1 - \theta(1 - \gamma))\tau$.

Before analyzing the optimal welfare-maximizing inflation rate, it is worth discussing the externalities surrounding the decision of an agent to join the financial system. When a buyer decides whether to join the financial system, she does not consider the consequences this has on the tax rate that the other buyers need to pay. For example, if at a certain tax level, a buyer decides not to join, this might increase the tax rate and induce more buyers to not join the financial system. In an extreme case, the "official" tax rate becomes very high so that no agent joins the financial system, and all suffer the costs of holding cash but pay only a share of the official tax rate. In this case, the government can raise inflation, thus making joining the financial system more attractive through three avenues: 1) the cost of staying out of the financial system increases, 2) the official tax rate decreases because the tax base widens, and 3) a lower share of government revenue is financed through taxation, reducing the tax base further. This has implications for the utility of the financially included and excluded and for aggregate welfare, as shown in Figure (3).

I show in the following that it is optimal to have positive inflation. In the case of the plentiful asset-bearing equilibrium, inserting the utilities $\mathcal{V}_i(x)$ and $\mathcal{V}_e$, welfare can be written as follows:

$$W = \int_0^{k\gamma} \Big((1-\rho)\mathcal{L} + \rho\mathcal{L}^* - (x + (1-\rho)c + \tau)\Big)\frac{1}{k}dx + (1-\gamma)(\mathcal{L} - c - (1-\theta)\tau)$$
$$= \gamma\Big((1-\rho)(\mathcal{L} - c) + \rho\mathcal{L}^* - \tau\Big) - \frac{1}{2}k\gamma^2 + (1-\gamma)(\mathcal{L} - c - (1-\theta)\tau)$$

The first-order condition w.r.t. μ can be written as follows:

$$\frac{\partial W}{\partial \mu} = \gamma\Big((1-\rho)\frac{\partial \mathcal{L}}{\partial \mu} - \frac{\partial \tau}{\partial \mu}\Big) + (1-\gamma)\Big(\frac{\partial \mathcal{L}}{\partial \mu} - (1-\theta)\frac{\partial \tau}{\partial \mu}\Big) = 0 \qquad (15)$$

The economic interpretation of equation (15) is that the first term represents the net welfare gain among the financially included, for whom welfare increases due to higher inflation (since $\frac{\partial \tau}{\partial \mu} < 0$), but the welfare from non-monitored meetings is lower. Note that the net benefit increases with ρ for the financially included.
The second term represents the welfare change for the financially excluded. On the one hand, welfare is lower because the financially excluded carry less cash in real terms and buy less in a DM meeting (since $\frac{\partial \mathcal{L}}{\partial \mu} < 0$); on the other hand, they profit from the reduction in taxes as well (as long as $\theta < 1$). The net benefit decreases with θ for the financially excluded.
It can be shown (see appendix for the proof) that[10]

$$\frac{\partial W}{\partial \mu}\Big|_{\mu=1} > 0$$

i.e., that it is optimal to have money growth that is higher than 1 and thus positive inflation.
Since it is not possible to find an analytic solution to the first-order condition (15), I provide a numerical analysis in the following section.

[10]The FOC can be alternatively written (noting that $\frac{\partial \gamma}{\partial \mu} = \frac{q}{k} - \frac{\theta}{k}\frac{\partial \tau}{\partial \mu}$) as:

$$\frac{\partial W}{\partial \mu} = -\frac{q}{\theta} + \frac{\partial \gamma}{\partial \mu}\frac{k}{\theta}(\gamma + (1-\gamma)(1-\theta)) = 0$$

4 Optimal Fiscal and Monetary Policy and Financial Inclusion

In this section, I discuss the optimal monetary and fiscal policies and the optimal level of financial inclusion under different technology parameters $\langle k, \theta \rangle$ which reflect the inefficiency in banking and tax collection. To do so, I consider specific functional forms and parameter values. The utility of the buyer in the DM is given by $u(q) = ln(q)$. The parameter values are summarized in table (1). The value for the discount rate β is set to the literature standard and G, the share of government spending in the goods consumed is set equal to the average value for central governments taken from the World Bank. The cost of using cash c is difficult to estimate. The value 2 percent of income is based on Malte and Seitz (2014), who provide an overview of these issues. The probability of meeting a financially included seller ρ is set to be approximately equal to the average financial inclusion of sellers and represents that of middle income countries. Finally, the ratio of currency to government liability δ can vary considerably. The value chosen ensures that the supply of bonds is abundant for all the relevant parameters k and θ.

Table 1: Parameter Values

Parameter	Description	Value
β	Discount factor	.95
G	Government spending (share of GDP)	0.2
c	Cost of using cash (share of income)	0.02
ρ	Prob. of meeting a financially included seller	0.5
δ	Ratio of currency to gov. liability	0.5
$F(\omega)$	Distribution of costs to access a bank	$U \sim [0, k]$, $k \in (0, 1)$
θ	Inefficiency in tax collection	$\theta \in (0, 1)$

The optimal inflation, tax revenue, and financial inclusion for different levels of

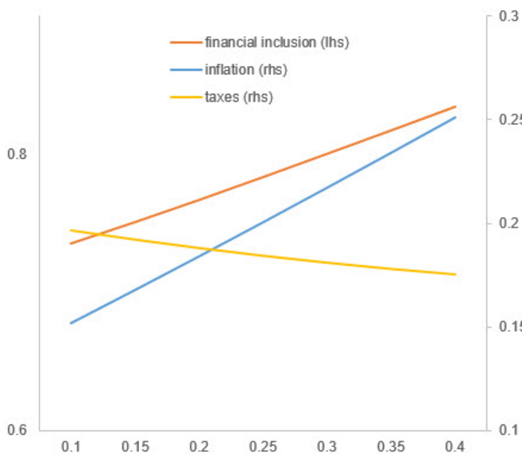
taxation and financial technology are displayed in figure (4). An improvement in taxation technology θ (leaving financial technology k fixed) leads to higher levels of financial inclusion, lower optimal inflation and higher tax revenue (see subfigure (4a)). Improving the taxation technology raises the tax revenue obtained from the financially excluded, which at the same time raises the attractiveness of joining the financial system and reduces the optimal inflation tax.
An improvement in the financial technology k (leaving taxation technology θ fixed) leads to a sharp increase in financial inclusion and tax revenue. Optimal inflation, however, increases slightly too (see subfigure (4b)); i.e., in such a case, financial inclusion and inflation are positively correlated. The reason is that with growing financial inclusion, the tax base increases, as does the optimal individual tax τ; thus the financially excluded also profit from a decrease in taxation. Therefore it is optimal to increase the inflation tax slightly.
Finally, if the financial and taxation technologies move jointly, then their improvement leads to a rapid increase in financial inclusion and in tax revenue. Optimal inflation also decreases sharply (see subfigure 4c).
Overall, in all three scenarios, technological improvements lead to higher financial inclusion and higher tax revenue; thus, both variables are always positively correlated. The correlation between inflation and financial inclusion is negative, whether taxation and financial technology move together or whether only taxation improves. In the case of an improvement in financial technology only, the correlation between financial inclusion and inflation is either positive or inconclusive. To test the model, I empirically analyze the relationship between these variables in the following sections.

5 Strategy for Empirical Analysis and Data

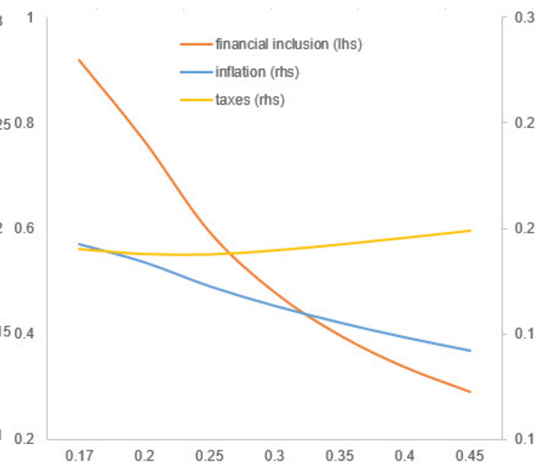
5.1 Model

In section 4, I have argued that tax revenue increases with financial inclusion and that depending on the technological progress, inflation decreases with financial inclusion. Therefore, regressing tax revenue and inflation on financial inclusion is the natural choice. However, the empirical analysis is made more difficult through

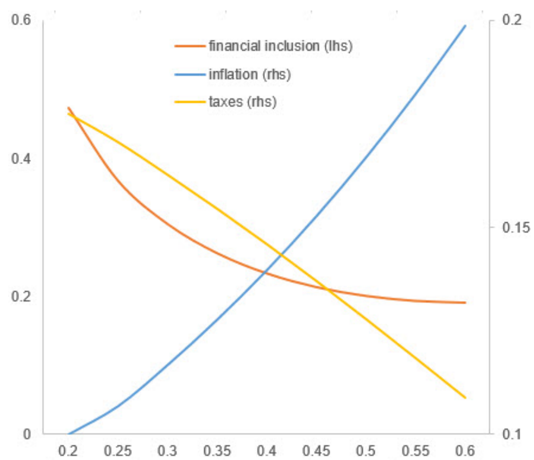
Figure 4: Taxes, inflation, and financial inclusion



(a) Taxation technology θ ($k = 0.2$ fixed)



(b) Financial technology k ($\theta = 0.2$ fixed)



(c) Technology $\theta = k$

two issues. First, the observable variables inflation, taxes, and financial inclusion $\langle \mu, \tau, \theta \rangle$ depend on the unobservable variables financial and taxation technology $\langle k, \omega \rangle$ and are set at the same time, blurring cause and effect. Second, while $\langle \mu, \tau \rangle$ can be observed for many countries over a long period, good quality data on γ for a broad range of countries are only available for the years 2011, 2014, and 2017, limiting the availability of data.

The strategy for the empirical analysis is to control for a broad set of macroeconomic variables that are known in the literature to affect inflation and tax revenue and to control for time and country-specific effects as well. More formally, I estimate the following model

$$\Gamma_{it} = \alpha\gamma_{it} + AX'_{it} + c_i + \lambda_t + \zeta_{it} \tag{16}$$

where the dependent variable Γ_{it} represents either inflation π_{it} or aggregate tax revenue T_{it}. γ_{it} measures financial inclusion (in terms of account ownership or in terms of transaction value or volume), and X'_{it} is a set of control variables that contains openness to international trade, government debt, the level of corruption, GDP growth, and unemployment. α and A represent the marginal impact of a change in financial inclusion or in the control variables on the dependent variables. c_i and λ_t represent country and year fixed effects, respectively. Finally, ζ_{it} is an error term.

5.2 The Datasets

I draw from two sources for the data on financial inclusion: the World Bank's Global Financial Inclusion survey, which looks at financial inclusion from the demand side, and the IMF's Financial Access survey, which covers the supply side. I describe each source in turn.

Since 2011, the Global Financial Inclusion Database has published a comprehensive data set on how adults save, borrow, make payments, and manage risks globally every three years.[11] The survey in 2017 covered 144 economies representing more than 97% of the world's population.[12] I have removed economies with only one observation or without at least two consecutive observations as well as all 19 economies that are part of the European Economic and Monetary Union because they do not control both their fiscal and monetary policies. For the years 2011, 2014, and 2017, the cleaned database contains 345 observations for 119 economies (thus, the panel data are not balanced due to missing observations for some economies).

In this paper, I focus on the payment aspects of financial inclusion and use the

[11] https://globalfindex.worldbank.org/

[12] For more details on the survey methodology, see Demirguc-Kunt et al. (2018), p. 111 ff.

variable FI, which measures the share of ownership of transaction accounts at a financial institution, as the main measure of financial inclusion (see table (2) for a set of descriptive statistics).[13] I check the robustness of this approach by using two different definitions of account ownership in the robustness section: FI+, which contains account ownership at financial institutions and mobile money institutes, and FI-, which measures usage of mobile money accounts only. The latter variable is available only for the years 2014 and 2017.

The Financial Access Survey[14] has been collecting data on the use of and access to basic financial services worldwide since 2004 and has covered account ownership and the usage of mobile money since 2007 in several countries (see IMF (2019) for the latest report). Due to the way the data are collected, the account ownership statistics are of limited value to this study because they include substantial double-counting of account owners. However, the statistics on the value of mobile money transactions (variable MTV), the number of mobile money transactions (variable MTN), the and number of active mobile money accounts (variable MAU) are very useful as a complement to my analysis (see table (3) for descriptive statistics).

5.3 Descriptive Statistics and Control Variables

Descriptive statistics of all dependent variables and control variables used in this paper are displayed in tables (2) and (3). The dependent variables tax revenue (Tax) and consumer price index (CPI) contain a fair share of variation across time (see within variation), but most of the variation occurs between countries. This statement also holds for most of the control variables.

Openness is defined as the ratio of imports and exports to GDP. There is no consensus in the literature on the impact of openness to trade on inflation and taxes. Romer (1993) argues that because openness increases the cost of inflation for economies, central banks have an incentive to reduce inflation with openness and provides empirical evidence in support of this claim. However, ample theoretical

[13] Financial institution accounts include those owned by respondents who report having an account at a bank or at another type of financial institution, such as a credit union, a microfinance institution, a cooperative, or the post office (if applicable), or having a debit card in their own name.

[14] https://data.imf.org/FAS

Table 2: Descriptive Statistics: World Bank Group Financial Inclusion Dataset

Variable	Definition	Panel	Mean	Sd	Min	Max	Obs.
Tax [a]	Tax revenue	Overall	15.40	5.99	0.07	36.50	N = 238
	(% of GDP)	Between		6.08	0.25	34.21	n = 90
		Within		1.03	11.46	19.41	T = 2.64
CPI [a,e]	Consumer prices	Overall	5.01	4.65	-3.70	29.51	N = 328
	(annual %)	Between		4.28	-1.01	26.09	n = 116
		Within		2.64	-2.02	17.97	T = 2.83
FI [b]	Acc. ownership at	Overall	0.48	0.30	0.02	1.00	N = 328
	fin. institutions	Between		0.29	0.05	1.00	n = 116
	(% age 15+)	Within		0.07	0.27	0.72	T = 2.832
FI+ [b]	Fin. institutions	Overall	0.49	0.29	0.02	1.00	N = 328
	and mobile money	Between		0.28	0.05	1.00	n = 116
	(% age 15+)	Within		0.08	0.26	0.73	T = 2.83
FI- [b]	Mobile money	Overall	0.10	0.13	0.00	0.73	N = 139
	(% age 15+)	Between		0.11	0.00	0.66	n = 76
		Within		0.06	-0.08	0.29	T = 1.83
Openness [a]	Trade	Overall	84.19	52.84	23.93	425.98	N = 326
	(% of GDP)	Between		51.37	24.25	408.07	n = 115
		Within		8.43	48.71	123.63	T = 2.83
Debt [a]	Central government	Overall	46.17	30.61	0.06	236.07	N = 319
	debt (% of GDP)	Between		29.37	0.23	231.05	n = 113
		Within		7.93	12.66	95.70	T = 2.82
COR [c]	Corrpution	Overall	41.92	18.18	8.00	92.00	N = 314
	Perception Index	Between		18.38	11.67	90.00	n = 111
		Within		2.03	32.58	48.42	T = 2.83
Δ GDP [a]	Yearly growth	Overall	16.72	17.73	0.68	96.55	N = 328
	GDP per capita	Between		17.72	0.76	86.40	n = 116
		Within		2.01	6.36	26.88	T = 2.83
UNP [d]	Unemployment	Overall	7.40	5.78	0.32	31.38	N = 325
	(% of labor force)	Between		5.70	0.44	27.26	n = 115
		Within		1.15	2.32	11.81	T = 2.83

Source: [a] World Bank; [b] Global Findex Database, World Bank; [c] Transparency International; [d] International Labour Organization.[e] Countries with CPI $\geq 30\%$ were removed from the sample

Table 3: Descriptive Statistics: IMF Financial Access Survey

Variable	Definition	Panel	Mean	Sd	Min	Max	Obs.
MTV	Value of mobile	Overall	7.93	17.20	0	142.39	N = 394
	money transactions	Between		10.01	0	41.03	n = 68
	(% of GDP)	Within		12.76	-33.1	119.84	T = 5.79
MTN	Number of mobile	Overall	7.25	15.99	0	195.97	N = 395
	money transactions	Between		8.45	0	39.06	n = 67
	(per person)	Within		12.78	-31.8	164.17	T =5.90
MAU	Number of active	Overall	0.15	0.20	0	0.94	N = 268
	mobile money accounts	Between		0.16	0	0.65	n = 50
	(per person)	Within		0.13	-0.27	0.85	T =5.36

Source: IMF Financial Access Survey

and empirical papers have either supported or contradicted the negative relation between openness and inflation (see Ghosh (2014) for a discussion). Gnangnon and Brun (2019) noted that economies that have successfully reformed their tax regimes generate higher tax revenues.

Central government debt (as a % of GDP), corruption, growth of GDP, and unemployment (as a % of the workforce) are the other major macroeconomic and institutional variables I control for.

6 Findings

Rapid and diverse changes in the level of financial inclusion in a relatively short period of time allow for an estimation of equation (16) with a fixed effects model.[15] The advantage of this approach is that the model controls for unobserved time-invariant heterogeneity between countries and that the estimators (mainly financial inclusion) can be endogenous with regard to the time-invariant heterogeneity without affecting the validity of the results. The fixed-effects model has the advantage

[15] The Hausman test clearly rejects the null hypothesis that the random effects model is the correct specification.

of controlling for such slow-moving institutional arrangements.

Table (4) presents all regression results. Columns 1 and 3 show a fixed-effects model with financial inclusion FI as the sole regressor. The log-transformed variable FI takes into account the fact that a 10 percentage point increase in financial inclusion in a country with a low level of basic inclusion is weighted higher than the same increase in a country with a high level. The coefficient on financial inclusion is positive (1.221) for tax revenues, negative (-4.057) for CPI and statistically significant in both cases, consistent with my model. The interpretation is that a doubling of the share of financially included adults in a country is associated with an increase in tax revenue of roughly 1.2 percentage points relative to GDP and a decrease in inflation of 4 percentage points.

Columns 2 and 4 show the same model with all control variables and country as well as year fixed effects. The FI coefficient w.r.t. taxes in Column 2 does not change much, while the coefficient in Column 4 decreases in magnitude (-2.291) but remains statistically significant. Openness and GDP growth are statistically significant for taxes (both positive) and government debt is significant in the CPI regression. The coefficients on corruption and unemployment are not statistically significant in either regression.

7 Robustness Checks

All results discussed in this section are summarized in tables (5) to (7).

7.1 Central Bank Independence

The theoretical model is based on implicit or explicit coordination between the fiscal and monetary policies. With an independent central bank, this coordination might break down. More precisely, the marginal effect of financial inclusion on inflation should be of lower magnitude (or not statistically significant) when accounting for central bank independence.

I use the central bank independence index from Garriga (2016) and divide the economies into two groups: independent central banks, which are at or above the median index value, and non-independent central banks, which are below the me-

Table 4: Estimation Results

Variables	(1) Tax	(2) Tax	(3) CPI	(4) CPI
FI (log)	1.221**	1.269**	-4.057***	-2.291***
	(0.477)	(0.490)	(0.666)	(0.860)
Openness		0.0370***		0.0349
		(0.0127)		(0.0248)
Debt		-0.0190		0.0604*
		(0.0188)		(0.0358)
COR		-0.0404		-0.0912
		(0.0347)		(0.0920)
Δ GDP		9.617**		-3.680
		(3.833)		(7.936)
UNP		-0.0252		0.101
		(0.0723)		(0.174)
Constant	16.36***	15.69***	0.848	1.575
	(0.406)	(2.107)	(0.671)	(6.350)
Observations	247	236	327	308
# of economies	92	88	116	109
Adj. R^2	0.054	0.173	0.175	0.257
Country FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	No	Yes
F-test	6.561	5.174	37.12	12.77

Standard errors (clustered at the country level) shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

dian.

Column 1 of table (5) reports the result for countries with independent central banks and column 2 for countries with non-independent central banks. The financial inclusion coefficient is statistically significant for independent central banks but not for that of the other group, although the magnitudes are similar. A careful interpretation of these results is that central bank independence does not play a significant role in the relationships observed.

Table 5: Estimation Results: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)
	Indep	Dep	Developing		Developed	
Variables	CPI	CPI	Tax	CPI	Tax	CPI
FI (log)	-2.666**	-2.143	1.152*	-2.686**	-1.495	0.745
	(1.173)	(1.458)	(0.603)	(1.118)	(1.083)	(1.155)
Openness	0.0317	0.0276	0.0579***	0.0444	0.0240	0.0317
	(0.0317)	(0.0409)	(0.0160)	(0.0534)	(0.0162)	(0.0224)
Debt	0.0733	0.0821*	-0.0608**	0.186***	0.00263	-0.0496
	(0.0572)	(0.0474)	(0.0239)	(0.0437)	(0.0274)	(0.0397)
COR	-0.232	0.0716	-0.0716	-0.000308	-0.0322	-0.191*
	(0.145)	(0.112)	(0.0640)	(0.112)	(0.0437)	(0.111)
Δ GDP	-4.896	-1.083	4.310	-5.152	6.324	0.466
	(14.54)	(9.350)	(9.192)	(17.71)	(4.567)	(7.372)
UNP	0.178	0.121		0.420	-0.104	0.0449
	(0.213)	(0.337)		(0.378)	(0.0761)	(0.163)
Constant	4.685	-5.459	13.89***	-9.264	15.61***	14.12*
	(7.979)	(9.250)	(3.055)	(8.214)	(2.440)	(8.173)
Observations	144	164	95	142	141	166
# of economies	50	59	37	51	51	58
Adj.R^2,Wald χ^2	0.389	0.163	0.428	0.371	0.112	0.259
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
F-test	9.171	7.610	6.234	11.13	1.934	7.302

Standard errors (clustered at the country level) shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

7.2 Developing vs. Developed Countries

I group low-income and low-middle-income countries into "Developing" countries and high-middle-income and high-income countries into "Developed" countries. For developing countries the coefficients on financial inclusion are significant and comparable in size to those from the basic regression for both tax and inflation (CPI). For developed countries the coefficient is not significant for either dependent variable.

Given that developed countries have a more developed taxation system, increasing financial inclusion does not affect tax revenue or inflation.

Table 6: Estimation Results: Different Measures of Financial Inclusion (World Bank)

Variables	(1) Tax	(2) CPI	(3) Tax	(4) CPI	(5) Tax	(6) CPI
FI (log)	1.269**	-2.291***				
	(0.490)	(0.860)				
FI+ (log)			1.192***	-1.796**		
			(0.380)	(0.742)		
FI- (log)					-0.290	-0.0412
					(0.255)	(0.490)
Constant	15.69***	1.575	15.56***	2.258	13.36***	1.768
	(2.107)	(6.350)	(2.090)	(6.429)	(4.162)	(8.300)
Observations	236	308	236	308	93	129
# of economies	88	109	88	109	54	71
Adj. R^2	0.173	0.257	0.182	0.250	0.111	0.011
F-test	5.174	12.77	5.436	13.18	1.455	1.991

Standard errors (clustered at the country level) shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Coefficients of all CV are suppressed, Country FE and Year FE: YES;

7.3 Other Measures of Financial Inclusion

Tables 6 and 7 estimate the impact on tax revenue and inflation using different measures of financial inclusion from the World Bank and the IMF dataset, respectively. The coefficients on the control variables are suppressed.

In table 6, I inserted Columns 2 and 4 from table (4) for comparison. Compared to these earlier estimates, the coefficients for FI+ in columns 3 and 4 are similar in size and statistically significant. However, the coefficients for FI- in columns 5 and 6 are not statistically significant. One possible reason for this surprising result is that FI-, which measures access to mobile money only, has been collected only in the years 2014 and 2017.

The estimation results presented in Table 7 use the measures of financial inclusion from the IMF dataset. In column 1, the coefficient on the value of mobile transactions (MTV) with regard to tax revenue is positive (0.0525) and statistically significant. The interpretation of the MTV coefficient is that an increase in the value of mobile phone transactions of 100 percentage points relative to GDP

Table 7: Estimation Results: Different Measures of Financial Inclusion (IMF)

Variables	(1) Tax	(2) CPI	(3) Tax	(4) CPI	(5) Tax	(6) CPI
MTV	0.0525**	-0.0146				
	(0.0245)	(0.0127)				
MTN			0.00958	-0.0526***		
			(0.0230)	(0.0150)		
MAU (log)					-0.123	-0.321*
					(0.110)	(0.176)
Constant	11.16*	-0.543	9.377	-3.110	7.935***	-2.968
	(5.692)	(3.349)	(5.702)	(2.524)	(2.322)	(3.363)
Observations	154	243	155	339	161	237
# of economies	44	60	44	61	38	46
Adj. R^2	0.175	0.132	0.124	0.196	0.175	0.122
F-test	2.706	3.474	2.031	6.606	3.344	3.323

Standard errors (clustered at the country level) shown in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Coefficients of all CV are suppressed, Country FE and Year FE: YES;

increases tax revenue by 5.25 percentage points. In column 2, the MTV coefficient is not statistically significant.

In column 4, the coefficient on the number of mobile transactions (MTN) with regard to inflation (CPI) is negative (-0.0526) and statistically significant. Accordingly, an increase of 100 mobile phone transactions per person reduces inflation by 5.26 percentage points. However, the coefficient in column 3 is not statistically significant (though it has the expected positive sign).

The last two columns, 5 and 6, show the results from regressing the number of active mobile money accounts per person on tax revenues and inflation (CPI). In this case, only column 6 has a statistically significant coefficient.

8 Concluding Remarks

This paper joins the discussion on the causes and the impact of financial inclusion by analyzing the link to technology and to monetary as well as fiscal policy.

I show that inefficient technologies, which lead to high costs of access to financial institutions, and an inefficient taxation system can lead to low levels of financial inclusion, low tax revenue and high (optimal) levels of inflation. Lowering the cost of access and improving taxation has the potential to not only raise financial inclusion (which in itself brings significant benefits for households and the economy as a whole) but also to increase the ability of governments to tax directly and lessens the pressure to do so through higher inflation.

The paper presents empirical evidence that this is already happening. One possible cause of the sudden and fast improvements in financial inclusion is arguably the rise of mobile payment services, cheaper identification, cheaper as well as faster payment infrastructure, and the appropriate regulation of these new payment providers.

The current empirical analysis does not provide a final and definitive answer on causality. I leave it to future research to address the issue of endogeneity and to find a suitable instrumental variable that correlates with financial inclusion but is not correlated with inflation and taxation.

Proof: Welfare Analysis

First Derivative

Consider the welfare function, which, after taking the integral, can be written as

$$W = \gamma\Big((1-\rho)\mathcal{L} + \rho\mathcal{L}^* - \tau\Big) - \frac{1}{2}k\gamma^2 + (1-\gamma)(\mathcal{L} - c - (1-\theta)\tau)$$

Taking the derivative w.r.t. μ, I obtain:

$$\frac{\partial W}{\partial \mu} = \frac{\partial \gamma}{\partial \mu}(1-\rho)(\mathcal{L}-c) + \gamma(1-\rho)\frac{\partial \mathcal{L}}{\partial \mu} + \frac{\partial \gamma}{\partial \mu}(\rho\mathcal{L}^* - \tau) - \gamma\frac{\partial \tau}{\partial \mu} - k\gamma\frac{\partial \gamma}{\partial \mu} - \frac{\partial \gamma}{\partial \mu}(\mathcal{L} - c - (1-\theta)\tau)$$
$$+(1-\gamma)(\frac{\partial \mathcal{L}}{\partial \mu} - (1-\theta)\frac{\partial \tau}{\partial \mu}) = 0$$

Rearranging, I obtain:

$$\frac{\partial W}{\partial \mu} = \frac{\partial \gamma}{\partial \mu}\underbrace{(\rho(\mathcal{L}^* - \mathcal{L} + c) - \theta\tau)}_{k\gamma} - \frac{\partial \gamma}{\partial \mu}k\gamma + \gamma((1-\rho)\frac{\partial \mathcal{L}}{\partial \mu} - \frac{\partial \tau}{\partial \mu}) + (1-\gamma)(\frac{\partial \mathcal{L}}{\partial \mu} - (1-\theta)\frac{\partial \tau}{\partial \mu}) = 0$$

Since the first two terms cancel each other out, I am left with equation (15).

First Derivative: Positive for $\mu = 1$

Based on equation (14), the derivative of τ w.r.t. μ can be expressed as:

$$\frac{\partial \tau}{\partial \mu} = -\frac{\partial \gamma}{\partial \mu}\Big(\frac{m_e\rho}{1-\theta+\theta\gamma} + \frac{m_e\theta(1-\gamma\rho)}{(1-\theta+\theta\gamma)^2}\Big)\Big(\frac{r-\delta r-1}{\delta} + \frac{1}{\mu}\Big) - \frac{m_e(1-\gamma\rho)}{1-\theta+\theta\gamma}\frac{1}{\mu^2} - \frac{\partial \gamma}{\partial \mu}\frac{\theta G}{(1-\theta+\theta\gamma)^2}$$

Re-arranging equation (15) to $\frac{\partial W}{\partial \mu} = \frac{\partial \mathcal{L}}{\partial \mu}(1-\gamma\rho) - \frac{\partial \tau}{\partial \mu}(1-\theta+\theta\gamma) = 0$ and inserting the above result as well as noting that $\frac{\partial \mathcal{L}}{\partial \mu} = -q = -\frac{m_e}{\mu}$, I obtain the following expression:

$$\frac{\partial W}{\partial \mu} = -(1-\rho\gamma)\frac{m_e}{\mu} + \frac{\partial \gamma}{\partial \mu}\Big(m_e\rho + \frac{m_e\theta(1-\gamma\rho)}{1-\theta+\theta\gamma}\Big)\Big(\frac{r-\delta r-1}{\delta} + \frac{1}{\mu}\Big) + (1-\gamma\rho)\frac{m_e}{\mu^2} + \frac{\partial \gamma}{\partial \mu}\frac{\theta G}{1-\theta+\theta\gamma}$$

Therefore, it follows that:

$$\frac{\partial W}{\partial \mu}\Big|_{\mu=1} = \frac{\partial \gamma}{\partial \mu}\theta\frac{G}{1-\theta(1-\gamma)} + \frac{\partial \gamma}{\partial \mu}\Big(m_e\rho + \frac{m_e\theta(1-\gamma\rho)}{1-\theta+\theta\gamma}\Big)\Big(\frac{r-\delta r-1}{\delta} + \frac{1}{\mu}\Big) > 0$$

References

Asli Demirguc-Kunt, Klapper Leora, Singer Dorothe, Ansar Saniya, and Hess Jake Richard. The global findex database 2017 : Measuring financial inclusion and the fintech revolution. *World Bank Group*, 2018.

Derryl D'Silva, Filkova Zuzana, Packer Frank, and Tiwari Siddharth. The design of digital financial infrastructure: lessons from india. *BIS Papers*, (106), 2019.

World Bank Group CPMI-IOSCO. Payment aspects of financial inclusion in the fintech era. techreport, Bank for International Settlements, World Bank Group, April 2020.

Richard M. Bird and Eric M. Zolt. Technology and taxation in developing countries: From hand to mouse. *National Tax Journal*, 61(4):791–821, 2008. ISSN 00280283, 19447477. URL `http://www.jstor.org/stable/41790481`.

Tatyana A. Koreshkova. A quantitative analysis of inflation as a tax on the underground economy. *Journal of Monetary Economics*, 53(4):773 – 796, 2006. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2005.02.009. URL `http://www.sciencedirect.com/science/article/pii/S0304393206000420`.

Juan Pablo Nicolini. Tax evasion and the optimal inflation tax. *Journal of Development Economics*, 55(1):215 – 232, 1998. ISSN 0304-3878. doi: https://doi.org/10.1016/S0304-3878(97)00063-1. URL `http://www.sciencedirect.com/science/article/pii/S0304387897000631`.

Ummad Mazhar and Pierre-Guillaume Meon. Taxing the unobservable: The impact of the shadow economy on inflation and taxation. *World Development*, 90:89 – 103, 2017. ISSN 0305-750X. doi: https://doi.org/10.1016/j.worlddev.2016.08.019. URL `http://www.sciencedirect.com/science/article/pii/S0305750X1630465X`.

Nouriel Roubini and Xavier Sala-i-Martin. A growth model of inflation, tax evasion, and financial repression. *Journal* *of Monetary Economics*, 35(2):275 – 301, 1995. ISSN 0304-3932. doi: https://doi.org/10.1016/0304-3932(95)01192-Q. URL `http://www.sciencedirect.com/science/article/pii/030439329501192Q`.

Ross Levine. Financial development and economic growth: Views and agenda. *Journal of Economic Literature*, 35(2):688–726, 1997. ISSN 00220515. URL `http://www.jstor.org/stable/2729790`.

Gamze Oz-Yalaman. Financial inclusion and tax revenue. *Central Bank Review*, 19(3):107 – 113, 2019. ISSN 1303-0701. doi: https://doi.org/10.1016/j.cbrev.2019.08.004. URL `http://www.sciencedirect.com/science/article/pii/S1303070119300794`.

Mohammed Ait Lahcen and Gomis-Porqueras Pedro. A model of endogenous financial inclusion: Implications for inequality and monetary policy. *University of Zurich Working Paper*, (310), 2019.

Morten Linnemann Bech and Jenny Hancock. Innovations in payments. *BIS Quarterly Review*, March 2020.

Thomas Philippon. On fintech and financial inclusion. *BIS Working Paper*, (841), February 2020.

Jon Frost, Gambacorta Leonardo, Huang Yi, Shin Hyun Song, and Zbinden Pablo. Bigtech and the changing structure of financial intermediation. *BIS Working Papers*, (779), April 2019.

Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113(3):463–484, 2005. ISSN 00223808, 1537534X. URL `http://www.jstor.org/stable/10.1086/429804`.

Guillaume Rocheteau and Randall Wright. Money in search equilibrium, in competitive equilibrium, and in competitive search equilibrium. *Econometrica*, 73(1):175–202, 2005. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/3598942`.

Ricardo Lagos, Guillaume Rocheteau, and Randall Wright. Liquidity: A new monetarist perspective. *Journal of Economic Literature*, 55(2):371–440, June 2017. doi: 10.1257/jel.20141195. URL http://www.aeaweb.org/articles?id=10.1257/jel.20141195.

Gabriele Berentsen, Aleksander Camera and Christopher Waller. Money, credit and banking. *Journal of Economic Theory*, 135(1):171 – 195, 2007. ISSN 0022-0531. doi: https://doi.org/10.1016/j.jet.2006.03.016. URL http://www.sciencedirect.com/science/article/pii/S0022053106001463.

Stephen D. Williamson. Liquidity, monetary policy, and the financial crisis: A new monetarist approach. *The American Economic Review*, 102(6):2570–2605, 2012. ISSN 00028282. URL http://www.jstor.org/stable/41724665.

Friedrich Schneider and Dominik H. Enste. Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 38(1):77–114, March 2000. doi: 10.1257/jel.38.1.77. URL https://www.aeaweb.org/articles?id=10.1257/jel.38.1.77.

Nathan Sussman. Debasements, royal reveneues, and inflation in france during the hundred years' war, 1415-1422. *The Journal of Economic History*, 53(1):44–70, 1993. ISSN 00220507, 14716372. URL http://www.jstor.org/stable/2123175.

Pedro Gomis-Porqueras, Adrian Peralta-Alva, and Christopher Waller. The shadow economy as an equilibrium outcome. *Journal of Economic Dynamics and Control*, 41:1 – 19, 2014. ISSN 0165-1889. doi: https://doi.org/10.1016/j.jedc.2014.02.006. URL http://www.sciencedirect.com/science/article/pii/S0165188914000517.

S. Boragan Aruoba. Institutions, tax evasion and optimal policy. *Working Paper.*, 2018.

Jane Ihrig and Karine S. Moe. Lurking in the shadows: the informal sector and government policy. *Journal of Development Economics*, 73(2):541 – 557, 2004. ISSN 0304-3878. doi: https://doi.org/10.1016/j.jdeveco.2003.04.004. URL http://www.sciencedirect.com/science/article/pii/S0304387803001688.

Mauricio Prado. Government policy in the formal and informal sectors. *European Economic Review*, 55(8):1120 – 1136, 2011. ISSN 0014-2921. doi: https://doi.org/10.1016/j.euroecorev.2011.04.010. URL `http://www.sciencedirect.com/science/article/pii/S001429211100050X`.

Satyajit Chatterjee and Dean Corbae. Endogenous market participation and the general equilibrium value of money. *Journal of Political Economy*, 100(3):615–646, 1992. ISSN 00223808, 1537534X. URL `http://www.jstor.org/stable/2138734`.

Douglas Gale Franklin Allen. Limited market participation and volatility of asset prices. *The American Economic Review*, 84(4):933–955, 1994. ISSN 00028282. URL `http://www.jstor.org/stable/2118039`.

Kruger Malte and Franz Seitz. Costs and benefits of cash and cashless payment instruments: overview and initial estimates. *Study commissioned by the Deutsche Bundesbank*, 2014.

IMF. Financial access survey 2019 trends and developments. techreport, International Monetary Fund, 2019.

David Romer. Openness and inflation: Theory and evidence. *The Quarterly Journal of Economics*, 108(4):869–903, 11 1993. ISSN 0033-5533. doi: 10.2307/2118453. URL `https://doi.org/10.2307/2118453`.

Amit Ghosh. How do openness and exchange-rate regimes affect inflation? *International Review of Economics & Finance*, 34:190 – 202, 2014. ISSN 1059-0560. doi: https://doi.org/10.1016/j.iref.2014.08.008. URL `http://www.sciencedirect.com/science/article/pii/S1059056014001105`.

S?na Kimm Gnangnon and Jean-Fran?ois Brun. Trade openness, tax reform and tax revenue in developing countries. *The World Economy*, 42(12):3515–3536, 2019. doi: 10.1111/twec.12858. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/twec.12858`.

Ana Garriga. Central Bank Independence in the World: A New Data Set, 2016. URL `https://doi.org/10.7910/DVN/I2BUGZ`.

Chapter 3

Central Counterparty Auctions and Loss Allocation

Central Counterparty Auctions and Loss Allocation*

Robert Oleschak
Swiss National Bank (SNB)
Bundesplatz 1, CH-3003 Bern
robert.oleschak@snb.ch

June 19, 2020

Abstract

I analyse first-price, single-item central counterparty (CCP) auctions triggered by a member's default. Bidding agents attach private values to the auctioned financial contracts and share potential losses with the CCP. The CCP chooses how losses are allocated and the time of auction to minimize its losses. I show that the loss allocation (e.g. juniorization) affects the expected revenue of the auction in a narrow sense but does not affect the size of losses. Recovery measures increase the safety and soundness of a CCP but can lead to undesirable delay of the auction. Tearing up contracts is the least cost-effective recovery measure.

Keywords: Central Counterparty, Default Management, Auctions, Recovery
JEL Classification: C72, D44, D53, D82, G23, G28

*I would like to thank Andreas Fuster, Thorsten V. Koeppl, Cyril Monnet, Rekha Oleschak, an anonymous referee, the seminar participants at the Bank for International Settlement (BIS), and the Swiss National Bank (SNB) for their insightful comments. Disclaimer: The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the author. They do not necessarily reflect the views of the Swiss National Bank (SNB). The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

1 Introduction

A central counterparty (CCP) reduces the propagation of shocks in a financial system by guaranteeing the performance of financial contracts (e.g., derivatives) between two parties, even when one party at some point is not able or willing to meet her obligations. If a default occurs, CCPs re-allocate all the contracts of the defaulting agent to the surviving agents. Auctions are used in cases where the position to be transferred is large in relation to market liquidity or where a central market does not exist. CCPs can use auctions to incentivise the surviving agents to provide higher bids than the current market price and, thus, avoid fire sale conditions, as demonstrated by Vuillemey [2019]. Potential losses that could occur as a result of the auction will typically be covered by the defaulting agent's collateral. Should the losses exceed the available collateral, the CCP and the surviving agents step in to cover any remaining losses based on a pre-defined loss allocation mechanism. An incentive that some CCPs started to implement recently is to juniorize the default fund contributions of those bidders who provided low or no bids at all (see CPMI-IOSCO [2019]).

To reduce systemic risks in the OTC derivatives markets, the G20 announced in 2009 that standardized OTC derivative contracts should be cleared through CCPs.[1] As a result, regulators throughout the world mandate the use of CCPs for OTC derivative contracts that are deemed sufficiently standardized. The CCP's market share in OTC derivatives markets has rapidly grown between 2008 and 2019, mainly for interest rate derivatives from 40% to 75% and for credit default swaps from 0% to 55%, according to BIS OTC derivatives statistics 2019.

Defaults at CCP are rare events, but in the last decade two notable defaults have occurred. The default of Lehman Brothers in 2008 has triggered auctions, liquidations and transfers at CCPs around the globe. For example LCH - a London-based CCP - auctioned, liquidated and transferred a $9 trillion interest rate swap portfolio (Umar et al. [2018]) and CME - a Chicago-based CCP - auctioned off sizeable energy, interest rate, and equities futures and options contracts together with

[1] "All standardised OTC derivative contracts should be traded on exchanges or electronic trading platforms, where appropriate, and cleared through central counterparties by end-2012 at the latest. OTC derivative contracts should be reported to trade repositories. Non-centrally cleared contracts should be subject to higher capital requirements."

$2 billion margin deposits (Valukas [2010]). Both CCPs successfully re-allocated Lehman's contracts without having to use their own or the surviving members' funds. However, in 2008 Einar Aas, a Norwegian trader, defaulted which triggered an auction of energy futures at Nasdaq OMX - a Swedish-based CCP - which lead to a loss of €7 million for the CCP, and €107 million for the surviving members (Umar et al. [2018]).

All auctions mentioned above have been conducted with a few selected bidders only without any loss arrangement geared towards incentivising bidders to provide higher bids. Even though CCPs have gained practical experience with auctions in the past, little is known about the theoretical properties of recently introduced incentives schemes under different loss scenarios.

In this paper, I will analyse a simple version of a juniorization scheme where the winner does not lose her default fund contribution (but the losers do) and combine it with the loss allocation mechanism of a CCP. In addition, I analyse the bidding strategy in case the default fund is exhausted by defining how the residual losses are distributed.

Research Question and Approach

What are the CCP's and the agents' incentives in an auction? Do the incentives depend on whether the default fund is exhausted or not? Can a CCP incentivise bidders to provide higher bids and how does this affect the losses carried by the CCP and the other non-defaulting agents? Are the bidders willing to participate in such auctions and continue meeting their obligations? This paper seeks to provide answers to these questions.

What makes auctions conducted by CCPs particularly interesting and different from the standard literature on auctions is its loss allocation; depending on the size of the loss, the bidders might have to carry some of the losses themselves, i.e., the payoff of an agent does not need to be zero if she fails to win the auction.

Analysing CCP auctions is important for two reasons. First, CCPs have gathered some knowledge over the years on how auctions work in less extreme cases[2] but

[2]For example, LCH and CME auctioned off Lehman's positions in 2008, or the auction at Nasdaq OMX in 2018 triggered by a default of a member.

have no or very little experience in situations where the default fund is entirely used up. Second, with the regulatory push to mandatory clearing, a broader group of agents now use CCPs. Auctions have been designed towards a smaller and more homogeneous group of participants. This is currently being rethought, including inviting clients to the auction or using the loss allocation to encourage bidding.[3] A consistent and clear framework supports the evaluation of possible choices.

The approach is best described by way of an example supported by Figure 1, which depicts time on the x-axis and the value of the financial contract of the defaulted agent on the y-axis. Assuming that an agent defaults at time D, the earliest date a CCP can conduct an auction is at $D+\eta$. As long as the CCP has not transferred the contract to another agent, she is subject to mark to market gains or losses. In the example provided, the value of the contract falls between default D and auction $A \geq D+\eta$, leading to a holding loss of $v_D - v_A$ at the time of the auction. In addition, during the auction, the winning bidder might provide a bid $\beta^* < v_A$, which is lower than the value of the contract leading to an additional auction loss. The overall loss $v_D - \beta^*$ will be allocated by the CCP according to pre-defined rules. Typically, the losses are first allocated to the defaulting agent (by using initial margins and default fund contribution). If that is not sufficient, then the CCP's share of equity (or skin in the game) is used. Finally, the CCP can use the surviving agents' remaining default fund and by applying additional recovery measures.

In this paper, I analyse the optimal bidding strategy given a loss allocation that covers all possible sizes of losses, investigate which loss allocation minimizes a CCP's expected loss, consider whether recovery measures (e.g., usage of initial margins, cash calls, variation margin haircuts and tear-ups) can cover any possible size of losses, and what the optimal time of auction A from the viewpoint of the CCP is. Finally, I analyse auctions with a subset of agents or where a tear-up (or termination) of contracts looms if the auction fails to provide a minimum prize.

[3] A recently published discussion paper CPMI-IOSCO [2019] aims at facilitating the sharing of existing practices and considerations that CCPs might want to take into account when designing an auction.

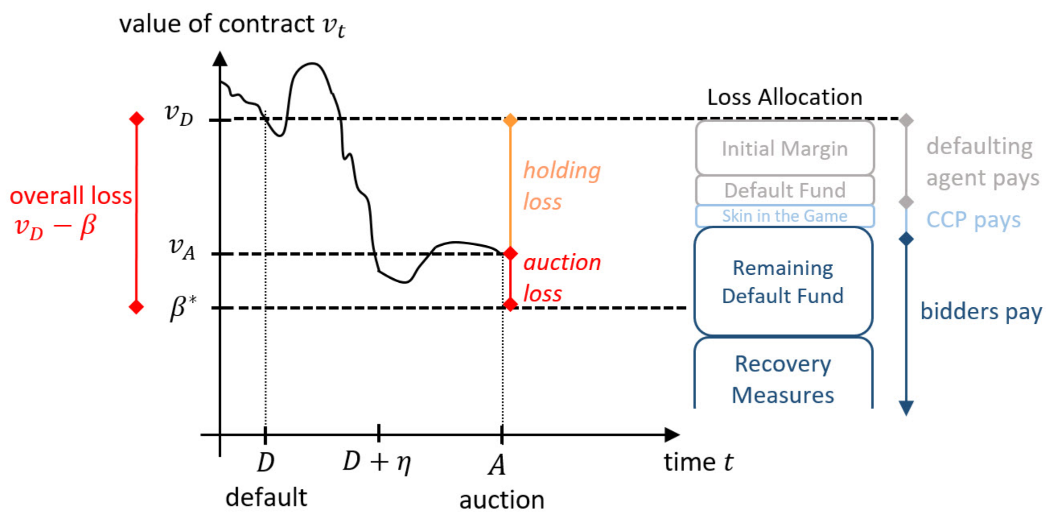


Figure 1: Auction Loss and Loss Allocation

Main Findings

First, I show that incentives can increase the auction revenue in a narrow sense but do no affect the payoffs of the CCP and the surviving agents. A CCP can increase the bids of the surviving agents by making the winner pay less for the auction loss. This incentive can be thought of as a simple juniorization of default fund contributions. The larger the loss that the winning agent can avoid, the higher the equilibrium price that the bidders submit. However, higher bidding does not translate into higher payoffs. This is true for both the CCP and the surviving agents because the gain of receiving higher bids is entirely consumed by the transfer to the winning bidder. Thus, the Revenue Equivalence Theorem holds if the transfers to the winner are considered.

Second, I show that the composition of agents in an auction can have a material impact on the outcome of the auction. Conducting auctions with a subset of agents has distributional effects because the agents invited to the auction are always better off than those who are not. The result crucially depends on how good the CCPs are at picking agents with high private values. Conversely, inviting additional agents to the auction (e.g., clients) could be inefficient (in the sense that the bidder with the highest private value does not need to win) but the CCP and

all losing bidders are better off when additional bidders are invited to the auction. Third, I show that recovery measures can affect the CCP's incentives in an auction. A CCP can design a loss allocation arrangement where the maximal loss is restricted to a share of her equity (typically required by regulators and called skin-in-the game) and that any losses beyond that can be allocated to the surviving agents (by using exhaustive recovery measures). The recovery measures can be designed in such a way that the surviving agents are not only willing to participate in an auction but, once the winning bid and losses are announced, are willing to meet all their obligations. Because the maximal loss of a CCP is capped, the CCP might in cases of high losses prefer to wait until the value of the contract has eventually rebounded rather than choosing to swiftly conduct an auction.
Finally, I show that tear-ups (partial or full) inflict higher costs on the financial system compared to other recovery measures, such as cash calls, variation margin haircuts or usage of initial margins. Thus, agents prefer to share the necessary costs in order to avoid a tear-up. However, the auction fails to coordinate the actions of the bidders, leading to an inferior equilibrium for all. Tear-ups should be avoided.

Relevant Literature

This paper draws on the well-established auction theory developed by many authors dating back to Vickrey [1961]. The literature on all-pay auctions (for example Baye et al. [1993] on interest group lobbying) resembles the problem of a CCP auction in that the losers might face negative payoffs. However, there is a crucial difference from the CCP auction in that the losing bidder's payment depends on the bids of others and not on her own bid.
Ferrara and Li [2017] apply the auction theory to CCPs. The authors analyse CCP auctions under static penalty schemes and note the inefficiencies that such an approach can have. In contemporaneous work Huang and Zhu [unpublished] study the incentives of bidders in a uniform-price CCP auction and find that juniorization of the guarantee fund contributions can elevate the equilibrium price.
The present paper extends the literature on CCP auctions in three ways. First, I study not only the incentives of the bidders in an auction but also those of the

CCP. This allows to identify potential moral hazard problems inherent in a CCP. I show that in case of extreme losses, a CCP might not act in the interest of her participants. Second, I consider auctions at all levels of losses, including where recovery measures become necessary. This allows to analyze the compatibility of different recovery measures with the auction format. A new result presented in this paper is that tear-ups do not incentivise the bidders to provide higher bids, even though all bidders would be willing to pay to avoid this recovery measure. Third, I explicitly formulate the CCPs budget as well as the auction participants' participation and termination constraint thereby extracting additional insights, including the interaction between participating and non-participation agents as well as the losses incurred on the whole financial system. An additional insight presented in this paper is the irrelevance of incentives, i.e. that incentives increase the size of the bids but do not affect the losses of the CCP or of the bidders. The notation used in this paper is based on Krishna [2010].

In the following section, I provide a short description of CCPs, with a focus on the loss allocation. In section 3, I characterize the model and state the CCP's and bidders' maximization problem subject to the CCP's loss allocation and formulate the constraints. In section 4, examples of loss allocation are discussed and the optimal bidding strategy is formulated. The general results are presented in three separate sections, as follows: the optimal loss allocation (section 5), the completeness of loss allocation and asymmetric auctions (section 6) and the optimal time of auction (section 7). Section 8 contains additional considerations, including auctions with a subset of agents and auctions against the background of a tear-up of contracts if the auction fails to provide a minimum prize. Section 9 closes with conclusions. The appendix contains all relevant proofs.

2 A Primer on Central Counterparties

A CCP is "an entity that interposes itself between counterparties to contracts traded in one or more financial markets, becoming the buyer to every seller and the seller to every buyer and thereby ensuring the performance of open contracts." (CPSS [2016]). A CCP consists of a set of rules and procedures defining i) how the liabilities arising from these contracts are measured and covered with the agent's collateral, ii) how these contracts will be re-allocated if one agent is unable (or unwilling) to fulfil them and iii) how losses will be allocated that might arise during the transfer of the positions to a new agent.[4] I will briefly discuss each element in turn.

Measures liabilities and requires collateral: The CCP continuously measures each agent's current and future liability arising from all contracts submitted for clearing. To limit and manage its counterparty risk, a CCP requires from all trading agents variation margin, initial margin and default fund contributions. The *variation margins* α is a cash payment from the agent with a positive liability to the agent with a negative liability. Variation margins, once paid, ensure that the current net liability of each agent is set to zero. The sizes of the variation margin payments cannot be predicted since they are a function of a stochastic market variable. In case an agent stops paying variation margins (and subsequently defaults), it might take time for the CCP to transfer the contract to another agent. Any losses resulting from changes in the valuation of the contract during the transfer period are covered by the other two types of collateral. The *initial margins* γ needs to be paid by both agents involved in the transaction once a contract is submitted to the CCP for clearing. The initial margins of a defaulting agent can be used to cover losses that might arise during the transfer of the positions to another agent.[5] Finally, the *default fund contribution* δ of each agent can be used to cover the losses arising from its own default but - as opposed to initial margins - from the default of another agent also. The default fund D is the sum of all default fund contributions $D = \sum_i \delta_i$.[6]

Re-allocates defaulter's positions: A CCP can re-allocate the defaulting agent's position by way of an auction to the highest bidder. To cover the losses that

[4] These sets of rules and procedures have evolved over more than a hundred years and differ substantially (see for example Kroszner [1999]). Recent regulatory initiatives (mainly the Principles for Financial Market Infrastructure CPMI-IOSCO [2012]) have, however, contributed to a homogenization of the CCP's rules and procedures.

[5] To be precise and connect it to the auction, the initial margins aim to cover losses that might arise in the period between market price v_D, i.e., when the defaulting participant paid its last variation margin, and any market price that might arise at the time of the auction v_A plus any auction losses with a certain degree of confidence. The modelling of the initial margins of a CCP is based on highly sophisticated risk models, which are not discussed in detail here.

[6] CCPs usually define the overall size of the default fund based on the overall level of risks that they assumed. The size of the default fund is usually based on some stress tests.

might arise as a result of the default and the following re-allocation, CCPs have defined rules regarding how losses will be allocated.

Order	**Who Pays What**	**Notation**	Auction View
1	Defaulting Agent's Initial Margins	γ	Somebody else pays
2	Defaulting Agent's Default Fund Contribution	δ	
3	CCP's Share of Equity (Skin in the Game)	ϵ	
4	Surviving Agents' Default Fund Contribution	$D - \delta$	Bidders pay
5	Surviving Agents' Committed Additional Resources (Recovery Measures) including Cash Calls, use of Initial Margins, and Variation Margin Haircuts	-	
6	CCP's Remaining Equity	$1 - \epsilon$	

Table 1: Loss Allocation of a CCP

Allocates Losses: There are three aspects to be considered. First, the CCP defines the order of who has to pay what (see Table 1), which is often coined as the default waterfall; typically the initial margin γ and default fund contribution δ of the defaulting agent and then the CCP's share of equity ϵ (or skin in the game) are used first. If that is not sufficient, then the surviving agents' default fund contributions and additional recovery measures are used. At the end of the waterfall, the CCP's remaining equity $1 - \epsilon$ will be used. Second, in cases where the surviving agents have to pay, the CCP can set up rules regarding how the losses are to be spread across the surviving agents and connect it to the outcome of an auction. For example, the CCP can equally share all losses or juniorise the default fund contribution by making the winner of the auction pay less compared the losers. Third, if the losses are covered by the first three layers, then from the viewpoint of a bidder, it is an auction where "somebody else pays" for the losses. It is only when the losses exceed $\gamma + \delta + \epsilon$ that the "bidders pay" auction point of view occurs. In such a case, the bidders can either lose their default fund contribution or have to commit additional resources, including cash calls, use of initial margins, or variation margin haircuts. These so called recovery measures are treated in detail in chapter 6.

3 The Model

The model consists of one continuous-time period, where $t \in [0, 1]$ depicts time. The economy is populated by $N+1$ risk-neutral agents holding two types of illiquid risky assets and one risk-neutral CCP. On the date $t = 0$, the agents can agree on a financial contract mandatorily cleared by a CCP. On the date $t = D$, one trader unexpectedly defaults due to an external valuation shock. The CCP can conduct an auction at time $A \in [D + \eta, 1)$, i.e., at the earliest, $D + \eta$ after default or at the latest, right before the expiration $t = 1$ to re-allocate the contract of the defaulted party. The losses will be shared based on a pre-agreed loss allocation arrangement. On the date 1, the payoffs of all risky assets are determined and, thus, all uncertainty is resolved.

3.1 Endowment and Cleared Market

On the date $t = 0$, the CCP is endowed with one unit of equity owned by third-parties and each agent is endowed with cash m and a risky asset. Let $B(t) = B_t$ follow some continuous-time stochastic martingale process where $B_0 = 0$ and $E_t[B_1] = B_t$. One half of the agents' (type 1) risky asset yields $\pi + B_1$ at $t = 1$, where π is a fixed return and B_1 is the final value of the continuous-time stochastic process B_t. The other half of the agents' (type 2) risky asset yields $\pi - B_1$, i.e., there is no aggregate risk because the aggregate supply of risk is zero.[7] Both types cannot dispose of the asset even if $\pi + B_t < 0$ for type 1 and $\pi - B_t < 0$ for type 2.

The agents carry private cost c_i of holding the risky part B_t of the asset, which can be interpreted as the cost of managing the market risk or capital costs.

[7]The risky asset can be interpreted as a project that the agent has invested in and cannot trade with a fixed return and a variable return, which depends on some stochastic market value, for example, the interest or exchange rate.

Each type 1 agent is randomly matched with another type 2 agent at a central market where they trade a contract and subsequently clear with a CCP.[8] I formalize this contract as follows: type 1 will pay B_1 to type 2 at time 1. This contract shields both agents involved in the trade from the market risk B_t and so the cost c_i can be avoided.

Since this contract is cleared, each agent i commits to the CCP to i) to pay the variation margin α_i throughout the lifetime of the contract,[9] ii) pay the initial margin γ and default fund contribution δ at $t = 0$, and iii) share any losses based on a pre-agreed loss allocation arrangement, as depicted in Table 1. For each cleared contract, the agents bear uniform clearing costs k, which can be interpreted as the likelihood of an agent losing her default fund contribution or having to meet other obligations. I take all parameters as given and do not provide for the optimal size of collateral or equity.[10]

An agent i will prefer to trade the contract on a centrally cleared market as long as $k \leq c_i$. To keep the analysis simple, I will set $k = 0$ and assume that the agent always has sufficient cash to meet any variation margin payment α_t at any time.

3.2 Default Management Auction Problem

At time $t = D$, a type 2 agent is hit by an exogenous negative valuation shock and, as a result, the agent stops paying the variation margins and subsequently defaults. The value of the contract at the time that the defaulting agent paid its last variation margin was $v_D = E_D[B_1] = B_D$. The CCP can conduct a single-item first-price auction at any time $A \in [D + \eta, 1)$ to re-allocate the contractual

[8]The agents could trade without a CCP. I assume that trading these contracts is subject to mandatory clearing.

[9]From the viewpoint of type 2, her current liability is $-B_t$; therefore, the value of the contract at time t from her point of view is $v_t = B_t$. To keep the current liability at zero at all times, the variation margin (again from the viewpoint of type 2) must be defined as $\alpha_t = -B_t$.

[10]Note, that the size of the initial margin γ and default fund contribution δ required for each contract cleared, CCP's equity and recovery measures that the CCP has at hand, as well as the order in which these financial resources would be used are given. However, even though I do not model the optimal size of the initial margin and default fund, there is an implicit trade-off that the CCP needs to consider. By increasing the required financial resources from her clearing agents, the CCP reduces the likelihood of her share of equity ϵ being used but increases the clearing cost k. Therefore, some agents would no longer find it profitable to trade the contract with the CCP if $k > c_i$.

obligations of the defaulting party to one of the remaining non-defaulting N agents. The value of the auctioned contract $v_A = E_A[B_1] = B_A$ is publicly observable. The private value $x_i = v_A - c_i$ that a bidder i attaches to the object at the time of the auction A is the value of the contract v_A minus the private costs c_i of holding the (additional) auctioned contractual obligation. The cost c_i re-appears because any agent holding an additional contract would not be optimally hedged any longer and would hold the risk B_t on her books. The distribution of the cost $c_i \sim U[0, 1]$ is i.i.d. across agents and is common knowledge. Therefore, from the viewpoint of the CCP, the private value of each bidder is uniformly distributed on the interval $x_i \sim U[v_A - 1, v_A]$. The corresponding cumulative distribution function is $F(x_i) = (x_i - (v_A - 1))$ with density $f(x_i) = 1$.

The CCP's Maximization Problem

Given that an agent has defaulted, the CCP conducts an auction to minimize the expected loss of equity L_{CCP} by choosing the optimal time of auction A, and loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot) \rangle$, as defined in equation (2), as follows:[11]

$$L_{CCP} = \begin{cases} 0 & \text{if } v_D - \beta(v_A - 1) \leq \gamma + \delta \\ \epsilon & \text{if } v_D - \beta(v_A) \geq \gamma + \delta + \epsilon \\ (0, \epsilon) & \text{if } x_i \in [v_A, v_{A-1}], \text{ where } \gamma + \delta < v_D - \beta(x_i) < \gamma + \delta + \epsilon \end{cases} \tag{1}$$

where $\beta(y)$ is the equilibrium bidding strategy of the winning bidder with private value y. The above equation simply states that the expected loss of a CCP conducting an auction at time A can have three states, as follows: either her share of equity will not be lost (first line), will definitely be lost (second line), or some of the equity will be lost (third line).

The Agent's Maximization Problem

I will now define the equilibrium bidding strategy β of the bidders. The bidder i providing bid b_i faces the following payoffs where $b_{-i} = \max b_{j \neq i}$ is the best bid of

[11] To simplify the analysis, I will generally assume that all agents participate in the auction. Section 8.1 discusses the case where not all agents are invited to the auction.

the other $N-1$ bidders:

$$u_i = \begin{cases} x_i - b_i - \mathcal{W}(b_i) & \text{if } b_i > b_{-i} \\ -\mathcal{L}(b_{-i}) & \text{if } b_i < b_{-i} \end{cases} \tag{2}$$

Note that ties where $b_i = b_{-i}$ will occur with zero probability, so I will ignore it. The payoff to the bidder i is $x_i - b_i - \mathcal{W}(b_i)$ if she wins and is $-\mathcal{L}(b_{-i})$ if she loses.[12] $\mathcal{W}(b_i)$ defines the loss or profit a CCP will inflict on bidder i with the bid b_i given that she has won. $\mathcal{L}(b_{-i})$ defines the loss or profit a CCP will inflict on bidder i given that bidder $-i$ with bid b_{-i} has won. The loss allocation as described in section 2, including all possible recovery measures, can be analysed in this framework.

Given that the other bidders follow the symmetric, increasing, and differentiable equilibrium strategy β, bidder i maximizes the following expected profit by choosing the optimal bid b_i as follows:

$$\pi(x_i, b_i) = G(\beta^{-1}(b_i))(x_i - b_i - \mathcal{W}(b_i)) - \int_{\beta^{-1}(b_i)}^{v_A} \mathcal{L}(\beta(y))dG(y) - d(A)\xi_i \tag{3}$$

where $G(y) = F(y)^{N-1}$ denotes the distribution of the highest order statistics of the remaining $N-1$ bidders that bidder i is competing against.[13] If the bidder wins, then her expected profit is given by the first expression. If she loses, then her expected loss is given by the integral of the loss function and the corresponding distribution of the highest bid. The last expression expresses costs $\xi_i \leq c_i$, which reflects the agents indirect exposure towards market risks should the CCP decide not to hold the auction at the earliest point $\bar{A} = \eta + D$ but to delay it. Therefore $d(A = \bar{A}) = 0$ and $d(A > \bar{A}) = 1$.

Forming the first-order condition and noting that in equilibrium $b_i = \beta(x_i)$, I

[12] In this paper, I use a special case of the loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot) \rangle$, where only my bid b_i and the highest bid of all the other bids b_{-i} define the loss allocation. A more general function would require as input all bids; thus, $\mathcal{W} : R^N \to R$. Such a function would be more difficult to manage since it would require the handling of the 2nd, 3rd,..., Nth-highest order statistics.

[13] For a detailed discussion of the order statistics, see Krishna [2010], Appendix C.

obtain the following:

$$G(x_i)\beta'(x_i)\Big(1+\mathcal{W}'(\beta(x_i))\Big)+g(x_i)\Big(\beta(x_i)+\mathcal{W}(\beta(x_i))-\mathcal{L}(\beta(x_i))\Big)=x_i g(x_i) \quad (4)$$

The equation above is the basis for calculating the equilibrium bidding strategy and the expected CCP's auction loss. Note that even though in (2) the payoff $\mathcal{L}(b_{-i})$ depends on the winning bid of somebody else, which cannot be known to the bidder in advance, the first-order condition only requires her own bid $\beta(x_i)$ as the input. Although no solution to this general form can be presented, I can note some general properties. Since β is increasing, and $G(x_i)$ as well as $g(x_i)$ are positive but fixed for agent i, there are two ways that the optimal bidding strategy can increase. First, if increasing the bid leads to a lower payment $\mathcal{W}'(\beta(x_i))<0$ and, second, if the losses if I lose are bigger than the losses if I win, i.e., $\mathcal{W}(\beta(x_i))-\mathcal{L}(\beta(x_i))<0$.

Constraints

There are *three types of constraints* that must be imposed on loss allocation $\langle\mathcal{W}(\cdot),\mathcal{L}(\cdot)\rangle$.

First, the overall loss has to be accurately covered. In the case where the surviving agents are paying at the margin, the *budget constraint* can be written for any winning bid β^* as follows:

$$\underbrace{(v_D-\beta^*)}_{\text{loss}}-\underbrace{(\gamma+\delta+\epsilon)}_{\text{used collateral}}=\underbrace{(N-1)\mathcal{L}(\beta^*)+\mathcal{W}(\beta^*)}_{\text{loss allocation}} \quad (5)$$

Second, the CCP cannot force a bidder to participate in an auction that offers her less expected utility with the optimal bidding strategy β than when not participating. The *participation constraint* is satisfied if the expected profit of participating is at least as large as the expected profit of staying out of the auction, as follows:

$$\pi(x_i)\geq-\underbrace{\int_{v_A-1}^{v_A}\mathcal{L}(\beta(y))dG(y),}_{\text{expected loss when not participating}}\quad\forall i \quad (6)$$

Third, the agent could - once the outcome of the auction has been determined - decide to leave the CCP for good to avoid sharing in the losses. A CCP does not allow agents to terminate contracts during a default and has contractual powers to inflict losses l on any leaving agent. The CCP could return the agent's collateral γ and δ only after subtracting any losses that the agent contractually agreed to share or she could refuse to pay the variation margins α to the leaving agent. In addition, the leaving agent would suffer private costs $c_i \geq 0$. Therefore, to guarantee that ex post no bidders prefer to leave the CCP after any winning bid $\beta^*(x_i)$ has been announced, the following *termination constraint* must be satisfied:

$$\mathcal{L}(\beta^*(x_i)) \leq c_{-i} + l, \text{ where } c_{-i} = \min c_{j \neq i} \tag{7}$$

The exact form of l will depend on the losses a CCP can inflict on a leaving agent and will be defined in chapter 6.

Definition 1 *A loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot) \rangle$ that satisfies all three constraints in (5), (6), and (7) for all possible sizes of holding losses $v_D - v_A$ is complete.*

3.3 Efficient Allocation

Assuming that the planner's objective is to maximize the unweighted sum of agents' and the CCP's utilities, the obvious solution to the social planners problem is to allocate the position to the agent with the highest private value $x_h = \max\, c_i$ immediately after the default has occurred. Any delay in auctioning off the position inflicts costs on the financial system $\sum_i \xi_i$. An inefficient auction where the position is allocated to an agent with a lower private value x_l reduces aggregated utility by $x_h - x_l$.

4 Optimal Bidding: Examples

I now turn to the detailed analysis of the bidder's behaviour and consider the case where somebody else pays (section 4.1), as well as two examples of a loss allocation when the bidders pay at the margin (section 4.2). Additionally, I analyse situations where it is ex ante not clear whether bidders will have to pay or not (section 4.3). In the last section 4.4, I compare all loss allocation regimes. In all cases, a type 2 agent defaults. Finally, I will use the following definition for an optimal bidding strategy.

Definition 2 *The optimal bidding strategy $\beta(x_i)$ for a bidder with private value x_i is such that it maximizes (3) given the time of the auction A, the number of bidders N, and the loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot) \rangle$ subject to budget condition (5).*

The participation and termination constraints will not be considered at this point. A full treatment of all conditions, including their completeness, is provided in chapter 6.

4.1 Somebody Else Pays

The auction starts with the observation that the collateral of the defaulting agent and the CCP's share of equity are sufficient to cover the losses resulting from even the lowest possible bid:

$$v_D - \beta(v_A - 1) \leq \gamma + \delta + \epsilon$$

This includes the scenario where the position of the defaulting agent carries a profit. Bidder i's payoffs are expressed in equation (2), where the bidders do not share any losses, so $\mathcal{W}(\cdot) = \mathcal{L}(\cdot) = 0$. The first-order condition

$$G(x)\beta'(x) = g(x)(x - \beta(x))$$

states that the expected marginal cost of increasing the bid (lhs) must equal the marginal profit (rhs). The optimal bidding strategy can be expressed as follows:

$$\beta(x_i) = \frac{N-1}{N} x_i + \frac{1}{N}(v_A - 1) \tag{8}$$

The bidders provide quotes that are lower than their private values. The bids approach their private value as the number of bidders N increases.
The expected profit of a bidder with private value x_i is as follows:

$$\pi(x_i) = \frac{(x_i - (v_A - 1))^N}{N} \geq 0$$

Inviting more agents to the auction increases optimal bidding and reduces the expected profit of the bidders. Finally, the expected loss of a CCP is zero $L_{CCP} = 0$.
The result is very standard. I will use this bidding strategy as a benchmark.

4.2 Losses Covered by Bidders

In case the value of the position v_A is such that the best possible bid exhausts the defaulter's collateral and the CCP's share of equity, as follows:

$$v_D - \beta(v_A) > \gamma + \delta + \epsilon \tag{9}$$

then the bidders will share the losses at the margin and the expected loss of a CCP is her full share of equity $L_{CCP} = \epsilon$. In the following, I will discuss two loss allocation arrangements subject to budget condition (5).

Example 1: Equal Loss Allocation

The equal loss allocation shares the losses equally among all surviving agents, i.e., for any winning bid β_e^*, I have that $\mathcal{W}_e(\beta_e^*) = \mathcal{W}(\beta_e^*) = \mathcal{L}(\beta_e^*) > 0$. Given the budget constraint in (5), the loss allocation function can be expressed as $\mathcal{W}_e(\beta_e^*) = \frac{(v_D - \beta_e^*) - (\gamma + \delta)}{N}$. The first-order condition, which is as follows:

$$\frac{N-1}{N} G(x)\beta_e'(x) = g(x)(x - \beta_e(x)) \tag{10}$$

and the optimal bidding strategy

$$\beta_e(x) = \frac{N}{N+1}x + \frac{1}{N+1}(v_A - 1) \tag{11}$$

leads to higher bidding compared to the benchmark, which can be best explained by the first-order condition in (10). While the expected cost of increasing the bid has been lowered by the factor $\frac{N-1}{N} < 1$, the expected marginal profit remains the same.
Finally, the expected profit of a bidder is lower compared to the benchmark model,

as follows:

$$\pi_e(x_i) = \pi(x_i) - \frac{1}{N}\big((v_D - v_A) - (\gamma+\delta) + \frac{2}{N+1})\big)$$

Example 2: Losers Loss Allocation

This loss allocation shares losses equally among the losing bidders - the winning bidder does not share any losses, i.e., for any winning bid β_w^*, I have $\mathcal{W}_w(\beta_w^*) = 0$, and $\mathcal{L}_w(\beta_w^*) > 0$. Given the budget constraint, the losing bidders pay $\mathcal{L}_w(\beta_w^*) = \frac{(v_D-\beta_w^*)-(\gamma+\delta+\epsilon)}{N-1}$. Inserting this into the first-order condition (4), I obtain the following:

$$G(x)\beta_w'(x) = g(x)(x - \beta_w(x) + \mathcal{L}(\beta_w(x))) \tag{12}$$

leading to the optimal bidding strategy, as follows:

$$\beta_w(x_i) = \underbrace{\frac{N-1}{N+1}x_i + \frac{N-1}{N(N+1)}(v_A-1)}_{\frac{N-1}{N}\beta_e(x)} + \frac{v_D - (\gamma+\delta+\epsilon)}{N} \tag{13}$$

The expected profit of a bidder in the case that the losers' loss allocation has the same value as the profit of a bidder in the equal sharing function, i.e., $\pi_w(x_i) = \pi_e(x_i), \forall i$.

4.3 Overlapping Losses

Consider that there is a bid $\beta_e(\psi) = \sigma$ with $\psi \in (v_A - 1, v_A)$ where the defaulting agent's collateral as well as the CCP's equity share is used up, i.e.,

$$v_D - \sigma = \gamma + \delta + \epsilon$$

Given the equal loss allocation $\mathcal{W}_e(\cdot)$, bidders face the following payoff:

$$u_i = \begin{cases} x_i - b_i & \text{if } b_i > b_{-i} \wedge b_i \geq \sigma \\ 0 & \text{if } b_i < b_{-i} \wedge b_j \geq \sigma \\ x_i - b_i - \mathcal{W}_e(b_i) & \text{if } b_i > b_{-i} \wedge b_i < \sigma \\ -\mathcal{W}_e(b_{-i}) & \text{if } b_i < b_{-i} \wedge b_j < \sigma \end{cases} \tag{14}$$

which simply means that as long as the winning bid is equal to or higher than σ, the bidders do not pay (because the defaulter's collateral is sufficient) or else they pay $\mathcal{W}_e(\cdot)$. I will solve the problem by backward induction.
First, for those bidders whose optimal equilibrium bid lies below σ, only the last three lines in equation (14) apply. The optimal bid b maximizes the following expected profit (given that the other bidders follow some equilibrium strategy β_D):

$$\pi(x, b_i) = G(\beta_D^{-1}(b_i))(x_i - b_i - \mathcal{W}_e(b_i)) - \int_{\beta_D^{-1}(b_i)}^{\psi} \mathcal{W}_e(\beta_D(y))dG(y)$$

Since the bidder can only influence his own likelihood of winning and the lower bound of the integral, the first-order condition as well as the optimal bid are identical to equations (10) and (11).
Second, for any bidder whose bid is higher than or equal to σ, only the first two lines of the payoff function apply. Therefore, she needs to find the optimal bid to maximize the following profit:

$$\pi(x, b_i) = G(\beta_D^{-1}(b_i))(x - b_i) \tag{15}$$

To find the optimal bidding strategy, special attention needs to be paid to the transition from bids below and above σ. The following reasoning is supported by Figure (2); starting with the private value $v_A - 1$, the optimal bidding strategy follows $\beta_D = \beta_e$, as expressed in (11), until the private value $\psi = \beta_e^{-1}(\sigma)$ is reached, which constitutes the boundary between usage of the default fund or not. At this point, i.e., A in the figure, the bidder's loss allocation function drops to zero ($\mathcal{W}_e = 0$) and the optimal bidding-curve in such a case would be β, as expressed in (8). However, jumping to point B is not an option since the bidder would re-enter the range where the loss allocation function is not zero and she would immediately jump back to the bidding curve β_e and so forth. Moving horizontally to point C and then following the β bidding curve is not optimal either since any bidder on this vertical line could increase her likelihood of winning by increasing her bid by a small amount and be better off. The solution is, therefore, a convergence to the

β bidding curve, as expressed by the following equation:

$$\beta_D(x_i) = \begin{cases} \frac{N-1}{N} x_i + \frac{1}{N}(v_A - 1) + \kappa & \text{if } \beta_D \geq \sigma \\ \beta_e(x_i) & \text{if } \beta_D < \sigma \end{cases} \tag{16}$$

where $\kappa = \frac{1}{N(N+1)} \frac{(\psi-(v_A-1))^N}{(x-(v_A-1))^{N-1}}$.

The implication is that all bidders, even those who provide bids that will not use the default fund, increase their bids.

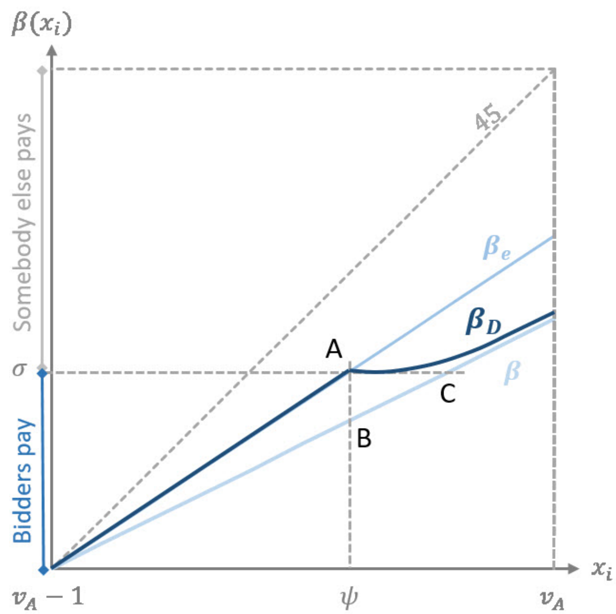


Figure 2: Optimal Bidding β_D when Bidders Might Have to Pay

Finally, it is easy to show that the bidding curve β_w must intersect the bidding curve β_e at point A, so that β_D for $\beta_D \geq \sigma$ follows the same path, independent of the loss allocation arrangement (as depicted on right panel in Figure 3). Take equation (13), where $\beta_w(\psi) = \frac{N-1}{N}\beta_e(\psi) + \frac{v_D-(\gamma+\delta+\epsilon)}{N}$, inserting $v_D - (\gamma + \delta + \epsilon) = \beta_e(\psi)$, I obtain $\beta_w(\psi) = \beta_e(\psi)$, which proves that bidding curves β_e and β_w intersect at $x_i = \psi$.

4.4 Comparison of the Loss Allocation Functions

A comparison of the optimal bidding strategies shows the following pattern.
First, the left panel of Figure 3 shows that for each private value $x_i \in [v_A - 1, v_A]$, I have that $\beta \leq \beta_e < \beta_w$, where β plots the optimal bidding curve in case the defaulting agent or the CCP covers the marginal costs (somebody else pays from the viewpoint of the bidders), β_e tracks the optimal bidding curve in case the default fund of the bidders is used but where the auction loss is shared equally among the bidders, and, finally, β_w is the optimal bidding where only the losing bidders cover the marginal auction losses. β_w leads to aggressive bidding of all parties, especially of those with low private values (or high costs c_i).
Second, the right panel of figure 3 shows that the above result is only valid as long as the bids lead to the bidders marginally sharing the loss, i.e., where $v_D - \beta > \gamma + \delta + \epsilon$. As soon as the bids enter the space where somebody else (the CCP) pays, then $\beta_w = \beta_e$, i.e., the bidders provide the same bid, independent of the loss allocation, which means that the CCP's expected $L_{CCP}(\beta_e) = L_{CCP}(\beta_w)$ in equation (1) is equal for both loss allocation arrangements.

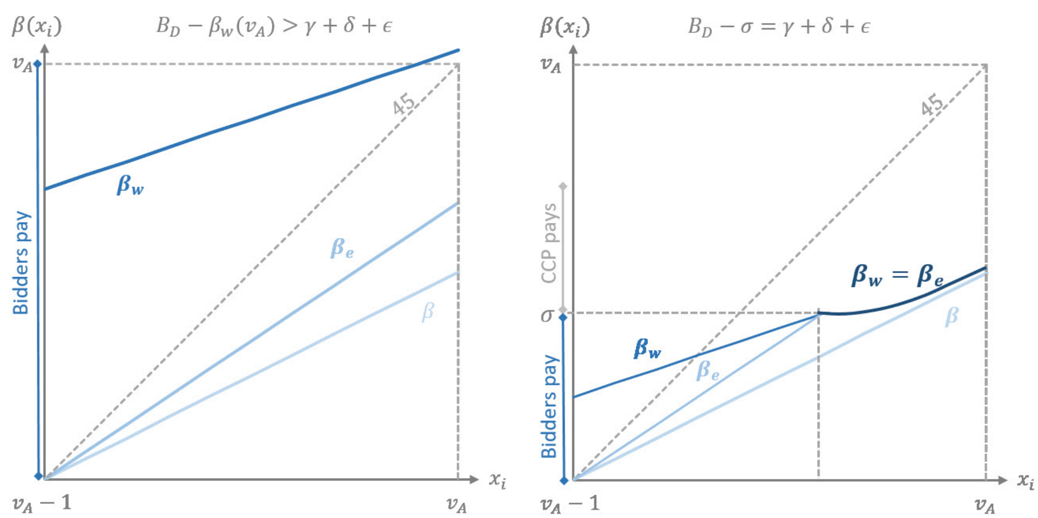


Figure 3: Equilibrium Bidding

Third, the expected profit of the bidders, as shown in the preceding sections, follows $\pi(x_i) > \pi_e(x_i) = \pi_w(x_i)$, i.e., the expected profit of the bidders is highest, in the case that somebody else pays. However, there is no difference in the expected profit for the equal and the losers' loss allocation.

5 Optimal Loss Allocation

In this chapter, I will analyse whether a CCP can minimize her expected loss by choosing an optimal loss allocation $\langle \mathcal{W}^*(\cdot), \mathcal{L}^*(\cdot)\rangle$, as defined below.

Definition 3 *The optimal loss allocation of a CCP $\langle \mathcal{W}^*(\cdot), \mathcal{L}^*(\cdot)\rangle$ when conducting an auction at time A with N number of participants, minimizes the expected loss of the CCP given by (1) subject to the budget condition (5).*

I can formulate the following two propositions (see proof in the appendix).

Proposition 1 *If the bidders might have to pay, then a bidder's as well as a CCP's expected (as well as ex ante) losses are independent of the CCP's loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot)\rangle$, i.e., any loss allocation is optimal.*

Intuitively, an agent optimizes $Z(x_i) = \beta(x_i) + \mathcal{W}(\beta(x_i))$ - the payment she would need to make to the CCP if she wins - which depends on exogenous factors and is not affected by the loss allocation $\mathcal{W}(\cdot)$. Given the optimal $Z(x_i)$, then $\mathcal{L}(x_i)$ can be directly derived and the expected profit of a bidder is shown to be independent of any loss-sharing arrangement, as follows:

$$\begin{aligned}
Z(x_i) &= \frac{N-1}{N+1}x_i + \frac{N-1}{N(N+1)}(v_A - 1) + \frac{v_D - (\gamma+\delta+\epsilon)}{N} \\
\mathcal{L}(x_i) &= -\frac{1}{N+1}x_i - \frac{1}{N(N+1)}(v_A - 1) + \frac{v_D - (\gamma+\delta+\epsilon)}{N} \\
\pi(x_i) &= \frac{(1-c_i)^N}{N} - \frac{1}{N}\big((v_D - v_A) - (\gamma+\delta+\epsilon) + \frac{2}{N+1}\big)
\end{aligned} \tag{17}$$

Proposition 2 *If the bidders might have to pay, then a CCP can incentivise higher bidding by reducing the loss allocation of the winner $\mathcal{W}(\cdot)$.*

The proof immediately follows from proposition 1. Since $Z(x_i)$ is independent of the loss allocation function, the optimal bidding strategy adjusts to the loss

allocation function, as follows: $\beta(x_i) = Z(x_i) - \mathcal{W}(\beta(x_i))$.
It is clear from the two propositions that by designing loss allocation $\langle \mathcal{W}^*(\cdot), \mathcal{L}^*(\cdot) \rangle$, a CCP can affect the optimal bidding strategy β but not the expected loss of the CCP nor of the bidder.

6 Completeness of Loss Allocation

A CCP can use recovery measures to allocate uncovered losses (see CPMI-IOSCO [2017] for a discussion). In this chapter I will consider cash calls, use of initial margins, and variation margin haircuts and analyse whether they are complete in the sense of Definition 1.
Proposition 1 simplifies the analysis in the following way. First, it can be shown that the PC from equation (6) is always satisfied when using the generally valid expressions from equation (17), which means that the PC is always satisfied for all types of loss allocation arrangements as long as the non-participating agent is subject to the same loss allocation arrangement as the participating bidders. Therefore, to verify completeness, I need to check only for the BC and the TC (see the proof in the appendix for proposition 4).
Second, for a given holding loss, the TC can be violated or not depending on the private value of the winning bidder and the bidder with the second highest bid. It can be shown (again based on the generally valid payments from equation (17)) that the TC is never violated whenever $\mathcal{L}(\beta_e(v_A)) \leq l$, i.e., I need to check only for the situation where the bidder i with private cost $c_i = 0$ wins and the second highest bidder j's private cost infinitesimally approaches zero, i.e., $\lim c_{-i} \to 0$ (see the proof in the appendix for proposition 4).
For ease of exposition, I will treat the default fund δ as already lost to the bidders. However, the CCP could organize an auction where she shields the winner from losing her default fund as well as from any additional recovery measures. As shown in proposition 2, this would increase the bidding dramatically, but based on proposition 1, would not affect the expected losses nor the ex post losses of the bidders or the CCP.

Cash Calls

This recovery measure covers additional losses by simply calling additional cash from the surviving agents so that given the winning bid β^*, the net loss is met by cash-calls $\mathcal{W}(\beta^*)$ from the winning agent and $\mathcal{L}(\beta^*)$ from the $N-1$ losing agents. In principle, the amount of cash that a CCP can call from the bidders is uncapped.[14] The budget constraint can be formulated as follows:

$$(v_D - \beta^*) - (\gamma + D + \epsilon) = \mathcal{W}(\beta^*) + (N-1)\mathcal{L}(\beta^*)$$

Since the default fund is already lost, the initial margin γ represents the maximum loss that a CCP can inflict on a losing bidder who wants to leave. For the TC to always be satisfied, I must have the following:

$$\mathcal{L}(\beta^*) \leq \gamma$$

The l.h.s depicts the cash call that a losing bidder would need to pay-in and the r.h.s depicts the maximal loss that a CCP can inflict on the leaving agent.

Use of Initial Margins

Additional losses are met simply by writing down the pre-paid initial margins, which the surviving agents would need to remargin again. It is obvious that the BC is violated whenever the loss is too big, i.e., as follows:

$$(v_D - \beta_e^*) - (\gamma + D + \epsilon) > N\gamma$$

The TC is met since prepaid resources are used to cover the losses.

Variation Margin Haircuts

This recovery tool achieves the budget balance by writing down the unrealized gains of in-the-money positions of the non-defaulting agents. Since I assume that a type 2 agent defaulted and losses occurred subsequently, all N_1 type 1 agents

[14] It has, however, become practice for the CCP to limit cash calls by a multiple of each agent's default fund contribution δ.

made gains and would be subject to vmgh. Since the loss-sharing applies only to a subset of agents, I first discuss the asymmetric payoffs, as follows:

$$u_i = \begin{cases} x_i - b_i - \delta - \mathcal{W}_1(b_i) & \text{if } b_i > b_{-i} \\ -\delta - \mathcal{L}_1(b_{-i}) & \text{if } b_i < b_{-i} \end{cases} \tag{18}$$

where $\mathcal{W}_1(\cdot)$ and $\mathcal{L}_1(\cdot)$ refer to the vmgh applied to type 1 agents only. This asymmetric loss allocation leads to a situation where type 1 agents bid more aggressively than type 2 members, i.e., $\beta_1(x_i) > \beta_2(x_i)$, but where both types provide higher bids compared to where one type bids on its own. Two results immediately follow. First, asymmetric auctions can be inefficient since for a small enough $\varepsilon > 0$, it is the case that $\beta_1(x_i - \varepsilon) > \beta_2(x_i + \varepsilon)$, so that the bidder with the higher private costs might obtain the position. Second, not allowing type 2 agents to participate ensures efficiency but reduces the competition between bidders and increases the CCP's expected loss. The results can be generalized as follows:

Proposition 3 *Auctions with asymmetric loss allocation can be inefficient. Bidders who are subject to the loss allocation bid more aggressively than those who are not. Adding latter can still increase the overall bidding, reduce the CCPs expected loss, and the losing bidders ex post loss.*

Continuing with the analysis of completeness, I will analyse for ease of exposition the equal loss allocation function and note that the maximum vmgh a CCP can apply to any type 1 agent is $v_D - v_A$.[15] The BC is always met whenever the following holds:

$$v_D - \beta_e^* < (\gamma + D + \epsilon) + N_1(v_D - v_A) \iff v_A - \beta_e^* < (\gamma + D + \epsilon) + (N_1 - 1)(v_D - v_A)$$

The difference $v_A - \beta_e^*$ is driven by the number of auction participants N and the private cost c_i^* of the winning bidder. The equation shows that the private cost of the winning bidder would need to be unrealistically large to violate this inequality. Basically, the difference between the market price and the winning bid would need to be large enough to wipe out all prepaid collateral as well as the holding loss

[15]This requires that the CCP stopped paying the variation margin to type 1 members immediately after the default.

multiplied by $(N_1 - 1)$.
Finally, the TC is always met since agent type 1 cannot dispose of the risky asset and therefore would be subject to the valuation loss $v_D - v_A$ anyway. In the following, I want to explore the TC in the case that the agents would be able to dispose of the risky asset at any time. The agent would keep the risky asset as long as the expected profit of the asset and the cleared hedged position is above zero, as follows:

$$\pi + \underbrace{(v_A - v_D)}_{\text{Loss due to holding risky asset}} - \underbrace{(\alpha_D - \alpha_A)}_{\text{Variation Margin Payments}} - \underbrace{\mathcal{L}_1(\beta^*)}_{\text{vmgh}} \geq 0$$

Note that $\alpha_D = v_D$ and $\alpha_A = v_A$, so that the vmgh cannot exceed $\mathcal{L}_1(\beta^*) \leq \pi$. In addition, if the agent leaves, the CCP can write down the initial margins γ also so that the TC is satisfied whenever the following holds:

$$\mathcal{L}_{1,e}(\beta^*) \leq \pi + \gamma$$

Combination of the Recovery Measures

Since the losses are allocated immediately, the initial margins γ can be counted only once. The CCP can either call cash up to the size of the γ or write down the initial margins up to γ, but it cannot do both. Combining either the cash call or the initial margin write down with the vmgh, the CCP can achieve a complete loss allocation. I obtain the following result.

Proposition 4 *Cash calls and initial margin write-downs are non-complementary recovery measures. A combination of the variation margin haircut and the write-down of initial margins or cash calls is complete as long as* $v_D - \beta_e^* < D + \epsilon + (N+1)\gamma + N_1(v_D - v_A)$.

As the following table shows, bringing de facto the initial margins into the CCP's loss allocation adds significant amounts of additional financial resources.
As shown, recovery measures enhance a CCP's ability to allocate losses considerably. I considered here only the willingness of the agents to continue meeting their obligations but not their ability to do so. The losses that the CCP allocates

	CME	LCH	ICE
D	7′416	10′205	2′472
ϵ	250	84	50
$(N+1)\gamma$	125′448	154′215	34′153

Table 2: Financial Resources of CCPs at Q42018, in USD Mio.

through the recovery measures described in this chapter affect the agents' capital and liquidity position almost immediately. The strength or vulnerability of a CCP is, therefore, her clearing agents, who, in one way or another, will have to carry the losses that go beyond the defaulting agent's collateral (and a small layer of the CCP's capital).

7 Optimal Time of Auction

I will analyse whether a CCP can minimize her expected loss by choosing the optimal time of the auction, as defined below.

Definition 4 *The optimal time of auction $A^* \in [D+\eta, 1)$ given a complete loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot) \rangle$ minimizes the expected loss of the CCP in (1).*

I will analyse the three cases of CCP's expected loss depicted in equation (1) at $\bar{A} = D + \eta$, i.e., the earliest time that a CCP can conduct an auction, as follows:

$$L_{CCP} = \begin{cases} 0 & \text{if } v_D - \beta(v_{\bar{A}} - 1) \leq \gamma + \delta \\ \epsilon & \text{if } v_D - \beta(v_{\bar{A}}) \geq \gamma + \delta + \epsilon \\ (0, \epsilon) & \text{if } x_i \in [v_{\bar{A}}, v_{\bar{A}} - 1] \text{ where } \gamma + \delta < v_D - \beta(x_i) < \gamma + \delta + \epsilon \end{cases}$$

When at $\bar{A}$, the CCP finds that $v_D - \beta(v_{\bar{A}} - 1) \leq \gamma + \delta$, i.e., she does not expect to lose her equity (first line), then the CCP will conduct the auction immediately. In this case, $A^* = D + \eta$.

However, if the CCP finds at $\bar{A}$ that $v_D - \beta(v_{\bar{A}}) > \gamma + \delta$, i.e., she expects to lose her share of equity for sure when conducting an auction right away (second line), then by waiting, she cannot lose more if the price declines further but can gain if the price rebounds. I do not calculate the optimal time A^* explicitly but note that the corresponding value v_{A^*} must satisfy in any case the following two inequalities at the same time: $v_D - \beta(v_{A^*}) < \gamma + \delta + \epsilon$ as well as $v_D - \beta(v_{A^*} - 1) > \gamma + \delta$. Since it is not guaranteed that the price will rebound, the CCP will wait to hold the auction until the price rebounds to value v_{A^*} or hold the position until $t = 1$. If the CCP finds at $\bar{A}$ that the expected loss of her equity is neither zero nor her full share ϵ, then she might decide to wait or conduct the auction right away. No clear answer can be given in such a case.

Proposition 5 *A CCP will conduct an auction at the earliest time possible $\bar{A} = D + \eta$ when the corresponding value of the auctioned contract $v_{\bar{A}}$ translates into no expected loss of equity. If the value of the contract at $\bar{A}$ is such that she will lose her entire share of equity ϵ for sure, then she prefers to wait until the price either rebounds or hold the position until maturity.*

The implication of a CCP holding the position for an extended period of time for the surviving members is that all agents are indirectly exposed to the market risk B_t and thus will carry cost ξ_i as expressed in equation (3). It can be argued that the type 1 members are more heavily exposed than type 2 members since the former would potentially be subject to *vmgh* in addition to potentially losing the initial margin γ. The CCP will not internalize these costs because all participants still weakly prefer to stay with the CCP. Crucially, this holds only for markets subject to mandatory clearing. If mandatory clearing was not in place, then two agents of opposite type could meet and agree to trade out of a CCP and hedge their exposure towards B_t in a bilateral contract.[16]

[16] This operation would require the agents to agree on two simultaneous transactions, as follows: first, an opposite trade to the one described in section 3.2 would be submitted for clearing to the CCP and, second, the same trade as described in section 3.2 would be agreed bilaterally.

8 Additional Consideration

8.1 Auctions with a Subset of Bidders

CCPs have in the past conducted auctions with only a subset of agents.[17] Therefore, it is of practical relevance to analyse how this affects the invited and not invited agents to the auction. I will analyse the situation where N_A is the number of agents invited to the auction but all agents $N > N_A$ are subject to the loss allocation of the CCP. It can be shown (see the proof in the appendix) that the payment of the winning bidder $Z_{N_A}(x_i)$ in such an auction is lower compared to an auction where all bidders are invited, which has direct implications for the non-invited agents who have to expect to share more losses. From this, the following proposition follows.

Proposition 6 *Inviting a subset of agents $N_A < N$ to the auction puts the non-invited agents at a comparative disadvantage.*

The benefits of increasing the number of agents participating in an auction crucially depend on how good the CCP is at picking agents with low private costs (or with a high private value). In this model, the CCP picks agents randomly. This is in contrast to the situation where the CCP does not know the exact private costs c_i but is able to divide all agents into two groups, as follows: agents with low private costs $c_i^l \sim U[0, x]$ and agents with high private costs $c_i^h \sim U[y, 1]$. In the simple case, where the support of the two groups does not intersect, i.e., $x \leq y$ (and where there are at least two low private cost agents), it can be shown that inviting the high cost agents to the auction will never improve the result and that the high cost agents will be indifferent regarding whether to participate in the auction or not.

The analysis can become very complicated when the support of the agents with low and high private costs intersect. Vickrey [1961] already noted that asymmetries among the bidders could lead to inefficient second-price auctions but noted that the mathematics of this problem might become intractable. Since then, many authors have tried to approach the problem from different angles. In addition to others,

[17] See, for example, a very well documented case of Lehman's Default at CME in Valukas [2010], p. 1841-1870.

Maskin and Riley [2000] have studied the properties of first-price and second-price auctions in the case of two asymmetric bidders whose stochastic values do not necessarily share the same support. Only recently did Hubbard and Kirkegaard [2015] extend the analysis to more than two bidders who do not have the same value support. They show that the results deviate crucially from the case of two bidders and extend the application to many additional relevant cases. Clearly, if the assumption of the CCP not knowing the private values of the agent is relaxed a bit, the results can change dramatically. Generally, it cannot be excluded that CCPs have some insights into the private values of the agents. After all, CCPs can observe the trading behaviour of the agents, their turnover, positions and so forth.

8.2 Re-Establishing a Matched Book: Partial Tear-Up

Another recovery measure a CCP can take is to terminate some or all contracts to return to a matched book and stem further losses (see CPMI-IOSCO [2017]). The CCP would establish a (market) price to calculate payments due to the affected surviving agents. If the available collateral is not sufficient, then the payments due would be reduced pro rata. In this paper, I will consider the partial tear-up, where the termination affects only those contracts necessary to offset the position of the defaulting agent. The results can be immediately translated into the termination of all the contracts of the CCP (complete tear-up).

The starting point is that the holding loss of the CCP is such that the available financial resources are not sufficient, i.e., that $v_D - v_A \geq \gamma + D + \epsilon$, and that the CCP would tear-up the contracts of the surviving agents unless the auction provides a sufficiently high winning bid β^*, where $v_D - \beta^* = \gamma + D + \epsilon$. Contracts once torn cannot be re-traded with another agent and the affected agent carries costs c_i since her risk is not fully hedged any longer. Is it desirable for the bidders to avoid a tear-up? And if yes, can an auction achieve a desirable outcome?

A partial tear-up is economically equivalent to the variation margin haircut, i.e., the CCP would partially tear-up a fraction $\frac{1}{N_1}$ of each agent type 1 contract with the CCP and the compensation would be reduced pro-rata as follows: $\mathcal{L}_1 =$

$\frac{(v_D - v_A) - (\gamma + D + \epsilon)}{N_1}$.[18] Given this, the CCP calls an auction and announces that positions will be partially torn and accordingly compensated if the winning bid β^* falls below $\beta^* < v_D - (\gamma + D + \epsilon)$. The bidders are faced with the following three possible outcomes and the respective payoffs (payoff in brackets apply to type 1 only):

$$u_i = \begin{cases} x_i - b_i - \delta & \text{if } b_i > b_{-i} \wedge v_D - b_i = \gamma + D + \epsilon \\ -\delta & \text{if } b_i < b_{-i} \wedge v_D - b_{-i} = \gamma + D + \epsilon \\ -\delta - (c_i + \mathcal{L}_1) & \text{if } v_D - b^* > \gamma + D + \epsilon \end{cases} \tag{19}$$

The first two payoffs describe the situation where the winning bid is sufficiently high to ensure that the CCP does not have to tear-up the positions. In the first line, the bidder i wins and in the second line, some other bidder $-i$ wins. The third line displays the payoffs in case the bids are not sufficiently high and the CCP would partially tear-up a fraction of each type 1's contract. Note that type 1 agent i still occurs costs c_i even though only a part of the contract was torn.
Do bidders have an incentive to provide bids as to avoid the tear-up?
To simplify the analysis, I consider the following reduced form in an example where two type 1 bidders i and j face an auction (bidders of type 2 are not considered here since they have no incentive to avoid a tear-up). They can either "pay" $b = v_D - (\gamma + D + \epsilon) = v_A + N_1\mathcal{L}_1$ and avoid the tear-up or "not pay" (meaning that they offer any price lower than b) and have their contracts torn. If both decide to pay, then the contract will be divided.

		Agent j	
		Pay	Not Pay
Agent i	Pay	$-c_i - \mathcal{L}_1,\ -c_j - h_1$	$-c_i - 2\mathcal{L}_1,\ 0$
	Not Pay	$0,\ -c_j - 2\mathcal{L}_1$	$-c_i - h_1,\ -c_j - \mathcal{L}_1$

Table 3: Payoff (δ not considered)

It is clear that all bidders prefer that someone else pays a high enough bid to avoid

[18] Each type 1 member would have a fraction $\frac{1}{N_1}$ of its contract torn and instead of receiving $\frac{v_A}{N_1}$ for the torn contract, it would receive only $\frac{v_D + (\gamma + D + \epsilon)}{N_1}$, so the loss would be $\frac{(v_D - v_A) - (\gamma + D + \epsilon)}{N_1}$.

the tear- up. The payoff in Table 3 shows that there can be only one equilibrium, i.e., both do not pay. Both would be equally well off if both paid or if none paid, but in the situation where both pay, each bidder has an incentive to deviate and not pay. Therefore, the bidders do not provide higher bids even if the partial tear-up is at risk.

The result of this game can be understood better if the variables are given concrete values. Consider the case where each type 1 would receive $\mathcal{L}_1 = -1$ less compensation for the share of torn contract and where $c_i = 0$ and $c_j = 1$ in table 4. From an aggregated point of view, the losses would be lowest if i pays and j does not pay since agent i has the lowest cost of storing another contract. In fact, agent j could compensate i and still be better off. Such an outcome could be achieved if the CCP would, instead of facing the bidders with a tear-up, consider covering the losses by the other recovery measures described in section 8 (e.g., cash-calls, initial margin haircuts).

		Agent j	
		Pay	Not Pay
Agent i	Pay	-1, -2	-2, 0
	Not Pay	0, -3	-1, -2

Table 4: Payoff with concrete values

In case of a full tear-up, all bidders would lose their positions, including those who were not affected by the default of the clearing member. The payoff in equation (19) now applies to all bidders (i.e., including the brackets), and it is easy to show that in this case, none of the bidders have an incentive to bid higher and avoid a full tear-up. The following proposition, therefore, holds for any type of tear-up.

Proposition 7 *Tear-ups are more costly on an aggregate basis compared to other recovery measures. Although agents would weakly prefer to share the losses instead of avoid having contracts torn-up, an auction cannot achieve the desirable outcome.*

9 Conclusions

In this paper, I formalize the incentives of a CCP and the surviving clearing agents in an auction conducted by the CCP when one agent defaults. I show that incentives (for example simple forms of juniorisation of default fund contributions) increase the overall bidding price but that this does not affect either the CCP's or the surviving agent's profits. In other words, even though clearing agents offer higher prices, less of their own default fund contributions will be deducted. The overall effect nets to zero.

A complete loss allocation can incentivise CCPs in certain instances to wait rather than quickly conduct an auction. This is for example the case in situations, where all size of losses are covered by the surviving agents by way of cash calls, use of initial margins, or variation margin haircuts. This can inflict costs on the financial system and has implications for the governance structure of a CCP. Far reaching decisions by a CCP, for example, the time of an auction, should incorporate the interests of the clearing agents.

Auction theory shows that having more agents participating in an auction increases competition and bidding prices. Conducting CCP auctions with a subset of agents has distributional implications too, where the invited agents are better off than those not invited to the auction. This applies, for example, to clients who might be subject to the loss allocation in the case of variation margin haircuts. Having more agents participate in an auction will not only raise the average winning bid but alleviate distributional issues.

Finally, I show, that tearing-up of contracts is an expensive recovery measure compared to other alternatives and that the threat of a tear-up does not coordinate agents sufficiently to bid higher prices to avoid the tear-up in the first place. CCP's should not rely on tear-ups as an incentive but use it only as a measure of last resort.

I have taken the size of the financial resources (e.g., default fund, initial margins, CCP's equity) as well as the default waterfall as given. Further research could build on this and frame the auction as a broader design problem that includes the optimal size of the financial resources provided by the clearing agents, the CCP as well as the order of the financial resources to be used.

Proofs

Optimal Bidding: Somebody Else Pays

Suppose that bidders $j \neq i$ follow the symmetric, increasing, and differentiable equilibrium bidding strategy β and that bidder i receives a signal that its private value is $x_i = v_A - c_i$ and bids b. In the following, I want to determine the optimal b.
The upper and lower limits of the optimal bid can be defined as follows. First, it is never optimal to bid $b > \beta(v_A)$ since bidder i would definitely win and can always do better by reducing its bid, and I will consider only bids $b \leq \beta(v_A)$. Second, a bidder with value $x = v_A - 1$ would never submit a bit that is higher than its private value since she would take a loss if she wins. If she bids lower, then she will definitely lose, and the profit is zero. Thus, it is a weakly dominant strategy to bid $\beta(v_A - 1) = v_A - 1$.
Bidder i wins the auction whenever she submits the highest bid $b_i > \max_{j \neq i} \beta(x_i)$. Since β is increasing, $\max_{j \neq i} \beta(x_i) = \beta(\max_{j \neq i} x_i)$ and, thus, bidder i wins whenever $b > \beta(\max_{j \neq i} x_i)$, or whenever $\beta^{-1}(b) > \max_{j \neq i} x_i$. The expected profit of a bidder is therefore as follows:

$$\pi(x, b) = G(\beta^{-1}(b))(x_i - b)$$

Maximizing w.r.t. b yields the following first-order condition where $G' = g$ is the density of the highest value, as follows:

$$\frac{g(\beta^{-1}(b)}{\beta'(\beta^{-1}(b))}(x - b) - G(\beta^{-1}(b)) = 0$$

At a symmetric equilibrium $b = \beta(x)$ and the equation above can be simplified as follows (the second expression is equivalent to the first one):

$$G(x)\beta'(x) + g(x)\beta(x) = xg(x)$$
$$\frac{d}{dx}(G(x)\beta(x)) = xg(x)$$

and since $G(v_A - 1) = 0$, we have the following:

$$\beta(x) = \frac{1}{G(x)} \int_{v_A - 1}^{x} yg(y)dy$$

Finally, by integrating by parts this can be rewritten as follows:

$$\beta(x) = x - \int_{v_A-1}^{x} \frac{G(y)}{G(x)} dy = \frac{N-1}{N} x_i + \frac{1}{N}(v_A - 1)$$

The first-order condition is a necessary but not sufficient condition. I refer to Krishna [2010] for the last part of the proof.

Optimal Bidding: Equal Loss Allocation

Inserting the risk-sharing function $f_e(\cdot) = f(\cdot) = h(\cdot)$ into equation (3) I obtain the following:

$$\pi_e(x_i, b) = G(\beta_e^{-1}(b))(x_i - b - f_e(b)) - \int_{\beta^{-1}(b)}^{v_A} f_e(\beta(y)) dG(y)$$

Maximizing w.r.t. b and by using the Leibnitz Rule for the integral yields the following first-order condition:

$$\frac{g(\beta_e^{-1}(b)}{\beta_e'(\beta_e^{-1}(b))}(x_i - b - f_e(b)) - G(\beta_e^{-1}(b))(1 + f_e'(b)) + \frac{g(\beta_e^{-1}(b)}{\beta_e'(\beta_e^{-1}(b))} f_e(b) = 0$$

At a symmetric equilibrium $b = \beta_e(x_i)$ the equation can be written as follows:

$$G(x_i)\beta'(x_i)(1 + f_e'(\beta_e(x_i))) + g(x_i)\beta(x) = x_i g(x_i)$$

After entering the respective values for $G(x_i) = (x_i - (v_A - 1))^{N-1}$ and $g(x_i) = (N - 1)(x_i - (v_A - 1))^{N-2}$, the above expression can be simplified into $(x_i - (v_A - 1))\beta'(x_i)\frac{1+f_e'}{N-1} + \beta(x_i) = x_i$, for which the solution is as follows:

$$\beta_e(x) = \frac{N-1}{N + f_e'} x + \frac{1 + f_e'}{N + f_e'}(v_A - 1)$$

There is no constant since it is weakly optimal for the bidder with private value $v_A - 1$ to bid its own value. For a concrete risk-sharing function $f_e' = -\frac{1}{N}$ I obtain the following:

$$\beta_e(x) = \frac{N}{N+1} x + \frac{1}{N+1}(v_A - 1)$$

Optimal Bidding: Losers Loss Allocation

Inserting the risk-sharing function $f_w(\cdot) = 0$ and $h_w(\cdot)$ into equation (3) I obtain the following:

$$\pi_w(x_i, b) = G(\beta_w^{-1}(b))(x_i - b) - \int_{\beta^{-1}(b)}^{v_A} h_w(\beta(y))dG(y)$$

Maximizing w.r.t. b and by using the Leibnitz Rule for the integral, I obtain the following first-order condition:

$$\frac{g(\beta^{-1}(b)}{\beta_w'(\beta_w^{-1}(b))}(x_i - b) - G(\beta_w^{-1}(b)) + \frac{g(\beta_w^{-1}(b)}{\beta_w'(\beta_w^{-1}(b))}h_w(b) = 0$$

At a symmetric equilibrium $b = \beta_w(x)$ and by simplifying $h_w = h_w(\beta_w(x))$, the above equation can be expressed as follows (the second line is a reformulation of the first):

$$\begin{aligned} G(x_i)\beta_w'(x_i) + g(x_i)(\beta_w(x_i) - h_w) &= x_i g(x_i) \\ \frac{d}{dx}G(x_i)(\beta_w(x_i) - h_w) &= x_i g(x_i) - G(x_i)h_w' \end{aligned}$$

Since $G(v_A - 1) = 0$, the second line can be formulated as follows:

$$G(x)(\beta_w(x) - h_w) = \int_{v_A-1}^{x} yg(y)dy - \int_{v_A-1}^{x} G(y)h_w'dy$$

Integrating by parts (where $\int uv' = [uv] - \int u'v$) and noting that $\int_{v_A-1}^{x} G(y)dy = \frac{1}{N}G(x_i)(x_i - (v_A - 1))$, as well as $h_w'' = 0$, I obtain the following:

$$G(x_i)(\beta_w(x_i) - h_w) = x_i G(x_i) - \frac{1}{N}G(x_i)(x_i - (v_A - 1)) - h_w'\frac{1}{N}G(x_i)(x_i - (v_A - 1))$$

Finally, by re-arranging the expression, I obtain the following:

$$\beta_w(x) = \frac{N - 1 - h_w'}{N}x + \frac{1 + h_w'}{N}(v_A - 1) + h_w$$

Note that this time, it is not weakly dominant for a bidder with private value $v_A - 1$ to bid its own value. Imagine a bidder with private value $v_A - 1 + \epsilon$, where ϵ is very small. If it wins, it does not have to pay h_w, but if it loses (which is very likely), if will have to pay h_w anyway. If it bids its own value, then another bidder close but with a slightly lower private value could increase its chances of winning by bidding slightly more than its own private value. Therefore, it cannot be an equilibrium.

For a concrete risk-sharing function $h_w(\beta_w(x_i)) = \frac{(v_D - \beta_w(x_i)) - (\gamma+\delta)}{N-1}$ I obtain the following:

$$\beta_w(x_i) = \frac{N-1}{N+1}x_i + \frac{N-1}{N(N+1)}(v_A - 1) + \frac{v_D - (\gamma+\delta)}{N}$$

Optimal Bidding: Overlapping Losses

The bidder whose bid is above the critical value $b \geq \sigma$ knows that the default fund will not be used and has to solve a maximization problem expressed by equation (15). The first-order condition can be written as follows:

$$\frac{d}{dx}(G(x)\beta_D(x)) = xg(x)$$

Since $\beta_D(\zeta) = \sigma$, I have that for any optimal bid above σ (i.e., where only the defaulter's collateral is used) is as follows:

$$\beta_D(x_i) = \sigma\frac{G(\zeta)}{G(x_i)} + \frac{1}{G(x)}\int_{\zeta}^{x} yg(y)dy$$

Integrating by parts yields the following:

$$\begin{aligned}\beta_D(x_i) &= \sigma\frac{G(\zeta)}{G(x)} + \frac{1}{G(x)}\Big([yG(y)]_{\zeta}^{x} - \int_{\zeta}^{x} G(y)dy\Big)\\ &= \sigma\frac{G(\zeta)}{G(x_i)} + x - \frac{G(\zeta)}{G(x_i)}\Big(\zeta - \frac{1}{N}(\zeta - (v_A - 1))\Big) - \frac{1}{N}(x_i - (v_A - 1))\\ \beta_D(x_i) &= \frac{N-1}{N}x_i + \frac{1}{N}(v_A - 1) + \frac{1}{N(N+1)}\frac{(\zeta - (v_A - 1))^N}{(x_i - (v_A - 1))^{N-1}}\end{aligned}$$

Comparison of the Loss Allocation Arrangements

Somebody Else Pays

The expected profit of a bidder with private value x_i in the case where someone else pays, i.e., with the optimal bidding strategy $\beta(x_i) = \frac{N-1}{N}x_i + \frac{1}{N}(v_A - 1)$ and where $l(x_i) = 0$ is as follows:

$$\begin{aligned}\pi(x_i) = G(x_i)(x_i - \beta(x_i)) &= (x_i - (v_A - 1))^{N-1}(x_i - \frac{N-1}{N}x_i - \frac{1}{N}(v_A - 1))\\ &= \frac{1}{N}(x_i - (v_A - 1))^N\end{aligned}$$

Equal Loss Allocation

The expected payoff of a bidder with private value x_i in the case where some bidders share the losses, i.e., with the optimal bidding strategy $\beta_e(x) = \frac{N}{N+1}x + \frac{1}{N+1}(v_A - 1)$ and the respective loss allocation function is:

$$\pi_e(x_i) = G(x_i)(x_i - \beta_e(x_i) - f_e(\beta_e(x_i))) - \int_{x_i}^{v_A} f_e(\beta_e(y))dG(y)$$

where $f_e(\beta_e(x_i)) = \frac{v_D - \beta_e(x_i) - (\gamma+\delta)}{N}$. The expression is more complicated, and I will calculate both parts separately. The first part can be calculated in two different ways, as follows:

$$\begin{aligned} &G(x_i)(x_i - \beta_e(x_i) - f_e(\beta_e(x_i)) \\ &= \frac{1}{N+1}\Big(x_i - (v_A - 1)\Big)^N - \Big(x_i - (v_A - 1)\Big)^{N-1}\frac{v_D - \beta_e(x) - (\gamma+\delta)}{N} \\ &= (x_i - (v_A - 1))^{N-1}\frac{1}{N}\Big(\frac{2N}{N+1}x_i - \frac{N-1}{(N+1)}(v_A + 1) - v_D + (\gamma+\delta)\Big) = \Upsilon \end{aligned}$$

The second part can be calculated in the following way:

$$\begin{aligned} &\int_{x_i}^{v_A} f_e(\beta_e(y))dG(y) = \int_{x_i}^{v_A} \frac{v_D - \beta_e(y) - (\gamma+\delta)}{N}dG(y) \\ &= \frac{N-1}{N}\int_{x_i}^{v_A}[v_D - (\gamma+\delta) - (\frac{N}{N+1}x_i + \frac{1}{N+1}(v_A - 1))](x_i - (v_A - 1))^{N-2} \\ &= \frac{N-1}{N}\Big[\Big(y - (v_A - 1)\Big)^{N-1}\Big(\frac{v_D - (\gamma+\delta)}{N-1} - 2\frac{v_A - 1}{N^2 - 1} - \frac{y}{N+1}\Big)\Big]_{x_i}^{v_A} \\ &= \frac{1}{N}\Big[\Big(y - (v_A - 1)\Big)^{N-1}\Big(v_D - (\gamma+\delta) - 2\frac{v_A - 1}{N+1} - y\frac{N-1}{N+1}\Big)\Big]_{x_i}^{v_A} \\ &= \frac{1}{N}\Big(v_D - (\gamma+\delta) - 2\frac{v_A - 1}{N+1} - v_A\frac{N-1}{N+1}\Big) - \Psi \\ &= \frac{1}{N}\Big((v_D - v_A) - (\gamma+\delta) + \frac{2}{N+1}\Big) - \Psi \end{aligned}$$

where $\Psi = \Big(x_i - (v_A - 1)\Big)^{N-1}\frac{1}{N}\Big(v_D - (\gamma+\delta) - 2\frac{v_A - 1}{N+1} - x_i\frac{N-1}{N+1}\Big)$. After combining both parts, I obtain the following result:

$$\Upsilon - \frac{1}{N}\Big((v_D - v_A) - (\gamma+\delta) + \frac{2}{N+1}\Big) + \Psi$$

where $\Upsilon + \Psi = \Big(x_i - (v_A - 1)\Big)^{N-1} \frac{1}{N}\Big(x_i(\frac{2N}{N+1} - \frac{N-1}{N+1}) - (v_A - 1)(\frac{2}{N+1} + \frac{N-1}{N+1}))\Big) = \frac{1}{N}\Big(x_i - (v_A - 1)\Big)^N$ so that finally the expected profit is as follows:

$$\pi_e(x_i) = \frac{1}{N}\Big(x_i - (v_A - 1)\Big)^N - \frac{1}{N}\Big((v_D - v_A) - (\gamma + \delta) + \frac{2}{N+1}\Big)$$

Losers Loss Allocation

The expected payoff of a bidder with private value x_i in the case where default fund contributions are seniorized and, thus, with the optimal bidding strategy $\beta_w(x) = \frac{N-1}{N+1}x + \frac{N-1}{N(N+1)}(v_A - 1) + \frac{v_D - (\gamma+\delta)}{N}$ is as follows:

$$\pi_w(x_i) = G(x_i)(x_i - \beta_w(x_i)) - \int_{x_i}^{v_A} h(\beta_w(y))dG(y) \tag{20}$$

where $h(\beta_w(y)) = \frac{v_D - \beta_w(y) - (\gamma+\delta)}{N-1}$. It can be shown that $\pi_w(x_i) = \pi_e(x_i), \forall x_i$, i.e., that the expected profit is the same for participants for both the equal sharing and the seniorization of the default fund.

Proof of Propositions 1 and 2

First, I will show that the expected as well as the ex-ante profit of the bidders is not affected by the loss allocation when bidders have to pay. In the second part, I will show that the CCP's expected profit is not affected either.

Independence of the Bidders' Profit

Define $Z(b_i) = b_i + f(b_i)$ as the payment that bidder i providing bid b_i would need to pay to the CCP if she wins. Then, her pay-off, as defined in equation (2), can be re-written as follows:

$$u_i = \begin{cases} x_i - Z(b_i) & \text{if } b_i > b_{-i} = \max b_{j \neq i} \\ -h(b_{-i}) & \text{if } b_i < b_{-i} \end{cases}$$

Since the budget constraint in equation (5) must be satisfied for all possible winning bids, I can define the payment that bidder i would need to make in case she loses as follows:

$$h(b_{-i}) = \frac{1}{N-1}\Big(v_D - (\gamma + \delta + \epsilon) - Z(b_{-i})\Big)$$

Therefore, the payoff to bidder i providing bid b_i can be expressed as follows:

$$u_i = \begin{cases} x_i - Z(b_i) & \text{if } b_i > b_{-i} = \max b_{j \neq i} \\ -\frac{1}{N-1}\Big(v_D - (\gamma + \delta + \epsilon) - Z(b_{-i})\Big) & \text{if } b_i < b_{-i} \end{cases}$$

Given that the other bidders follow the symmetric, increasing, and differentiable equilibrium strategy β, bidder i with the above payoff maximizes the following expected profit by choosing the optimal bid b_i, as follows:

$$\pi(x_i, b_i) = G(\beta^{-1}(b_i))(x_i - Z(b_i)) - \frac{1}{N-1}\int_{\beta^{-1}(b_i)}^{v_A} \Big(v_D - (\gamma + \delta + \epsilon) - Z(\beta(y))\Big) dG(y)$$

where $G(y) = F(y)^{N-1}$ denotes the distribution of the highest order statistics of the remaining $N-1$ bidders that bidder i is competing against.
Differentiating w.r.t. b_i (using the Leibnitz Rule for the integral) yields the following first-order condition, where $G' = g$ is the density of the highest order statistics (note also that $dG(\beta^{-1}(y)) = \frac{g(\beta^{-1}(y))}{\beta'(\beta^{-1}(y))}dy$), as follows:

$$\frac{g(\beta^{-1}(b_i))}{\beta'(\beta^{-1}(b_i))}(x_i - Z(b_i)) - G(\beta^{-1}(b_i))Z'(b_i) + \frac{1}{N-1}\frac{g(\beta^{-1}(b_i))}{\beta'(\beta^{-1}(b_i))}(v_D - (\gamma + \delta + \epsilon) - Z(b_i)) = 0$$

which can be simplified as follows (note, that $Z'(b_i) = (1 + f'(b_i))$:

$$\frac{N}{N-1}g(\beta^{-1}(b_i))Z(b_i)) + G(\beta^{-1}(b_i))\beta'(\beta^{-1}(b_i))(1 + f'(b_i)) = g(\beta^{-1}(b_i))(x_i + \frac{v_D - (\gamma + \delta + \epsilon)}{N-1})$$

At symmetric equilibrium $b_i = \beta(x_i)$, the first-order differential equation can be written as follows (note that $Z'(\beta(x_i)) = \beta'(x_i)(1 + f'(\beta(x_i))))$:

$$\frac{N}{N-1}g(x_i)Z(\beta(x_i))) + G(x_i)Z'(\beta(x_i)) = g(x_i)(x_i + \frac{v_D - (\gamma + \delta + \epsilon)}{N-1})$$

Since $Z(\beta(x_i))) = \beta(x_i) + f(\beta(x_i))$ only depends on x_i, I can replace it with $\hat{Z}(x_i) = Z(\beta(x_i)))$, as follows:

$$\frac{N}{N-1}g(x_i)\hat{Z}(x_i) + G(x_i)\hat{Z}'(x_i) = g(x_i)(x_i + \frac{v_D - (\gamma + \delta + \epsilon)}{N-1})$$

The solution to this first-order differential equation is as follows:

$$\hat{Z}(x_i) = c_1(v_A - 1 - x_i)^{-N} + \frac{N-1}{N+1}x_i + \frac{N-1}{N(N+1)}(v_A - 1) + \frac{v_D - (\gamma + \delta + \epsilon)}{N}, \text{ where } c_1 = 0$$

which proves that the payment that the winning bidder pays to the CCP is independent of the loss allocation arrangement.
Finally, the expected profit $\pi(x_i)$ as well as the ex post payment can be expressed in terms of $\hat{Z}(x_i)$, which finalizes the proof.

Independence of the CCP's Profit

Take the expected loss of a CCP as expressed in equation (1). The first two cases are not interesting since the expected loss is either zero or the complete share of equity ϵ will be lost. Therefore, I will show that in the third case, where $L_{CCP} \in (0, \epsilon)$, the loss allocation cannot affect the expected loss of a CCP.
Consider any possible loss allocation function $\langle \hat{\mathcal{W}}, \hat{\mathcal{L}} \rangle$ leading to the optimal bidding strategy $\hat{\beta}$, and where $\hat{\beta}(\zeta) = v_D - (\gamma + \delta + \epsilon)$ with $\zeta \in (v_A - 1, v_A)$, i.e., there is a bid that exactly exhausts all the defaulting agent's collateral as well as the CCP's share of equity. According to the contractual arrangements for any loss allocation measures, it must be that $\hat{\mathcal{W}}(\zeta) = \hat{\mathcal{L}}(\zeta) = 0$, and so inserting this into the above solution for $\hat{Z}(\zeta)$ leads to the following:

$$\hat{Z}(\zeta) = \hat{\beta}(\zeta) = \frac{N-1}{N+1}x_i + \frac{N-1}{N(N+1)}(v_A - 1) + \frac{\hat{\beta}(\zeta)}{N}$$

Solving for $\hat{\beta}(\zeta)$ I obtain the following:

$$\hat{\beta}(\zeta) = \frac{N}{N+1}\zeta + \frac{1}{N+1}(v_A - 1)$$

which is the solution for the optimal bidding strategy in the case of equal loss allocation, i.e., all loss allocation functions converge to point A in figure (2). Any bidding above $\hat{\beta}(\zeta)$ follows the same pattern, independent of the loss allocation function. From that, it follows that the expected loss L_{CCP} cannot be affected by the loss allocation function.

Proof of Proposition 3

First, I will show that including agents not subject to the loss allocation increases the optimal bidding strategy for both, including for those subject to the loss allocation. Second, I will show that this reduces the CCP's expected losses and, finally, I show that the losers' ex ante losses will also be lower as a result.

Adding agents not subject to the loss allocation

Add to the N agents subject to the loss allocation $\langle \mathcal{W}(\cdot), \mathcal{L}(\cdot)\rangle$ S agents who are not. Given winning bid β^*, the CCP's budget condition must satisfy one of two cases, as follows: either an agent N wins or where an agent S wins. In the case of an equal sharing function $\mathcal{W}_e(\cdot)$, the budget constraint will always be the same, as follows:

$$(v_D - \beta^*) - (\gamma + \delta + \epsilon) = N\mathcal{W}_e(\beta^*)$$

Given an equal sharing loss $W_e(\cdot)$, the first group of N agents choose b_N to maximize the following expected profit:

$$\begin{aligned}\pi_N(x_i, b_N) &= F(\beta_N^{-1}(b_N))^{(N-1)} F(\beta_S^{-1}(b_N))^S (x_i - b_N - \mathcal{W}_e(b_N)) \\ &- \int_{b_N}^{\beta_S(v_A)} \mathcal{W}_e(y) d\Big(F(\beta_N^{-1}(y))^{(N-1)} F(\beta_S^{-1}(y))^S\Big)\end{aligned}$$

where $F(y) = (y - (v_A - 1))$.
The second group of S agents maximize the following expected profit:

$$\pi_S(x_i, b_S) = F(\beta_N^{-1}(b_S))^N F(\beta_S^{-1}(b_S))^{(S-1)} (x_i - b_S)$$

The first-order condition of the N agent subject to loss allocation can be written as follows(second equation assumes that $b_N = \beta_N(x_i')$):

$$\left(\frac{(N-1)F(\beta_N^{-1}(b_N))^{(N-2)} F(\beta_S^{-1}(b_N))^S}{\beta_N'(\beta_N^{-1}(b_N))} + \frac{SF(\beta_N^{-1}(b_N))^{(N-1)} F(\beta_S^{-1}(b_N))^{(S-1)}}{\beta_S'(\beta_S^{-1}(b_N))}\right)(x_i' - b_N)$$
$$= F(\beta_N^{-1}(b_N))^{(N-1)} F(\beta_S^{-1}(b_N))^S (1 + \mathcal{W}_e'(b_N))$$
$$\left(\frac{(N-1)F(x_i')^{(N-2)} F(\beta_S^{-1}(\beta_N(x_i')))^S}{\beta_N'(x_i')} + \frac{SF(x_i')^{(N-1)} F(\beta_S^{-1}(\beta_N(x_i')))^{(S-1)}}{\beta_S'(\beta_S^{-1}(\beta_N(x_i')))}\right)(x_i' - \beta_N(x_i'))$$
$$= F(x_i')^{(N-1)} F(\beta_S^{-1}(\beta_N(x_i')))^S (1 + \mathcal{W}_e'(\beta_N(x_i')))$$

Additionally, the first-order condition of the S agents not subject to the loss allocation can be written as follows (second equation inserts $b_S = \beta_S(x_i)$):

$$\left(\frac{NF(\beta_N^{-1}(b_S))^{(N-1)}F(\beta_S^{-1}(b_S))^{(S-1)}}{\beta_N'(\beta_N^{-1}(b_S))} + \frac{(S-1)F(\beta_N^{-1}(b_S))^{N}F(\beta_S^{-1}(b_S))^{(S-2)}}{\beta_S'(\beta_S^{-1}(b_S))}\right)(x_i - b_S)$$
$$= F(\beta_N^{-1}(b_S))^{N}F(\beta_S^{-1}(b_S))^{(S-1)}$$
$$\left(\frac{NF(\beta_N^{-1}(\beta_S(x_i)))^{(N-1)}F(x_i)^{(S-1)}}{\beta_N'(\beta_N^{-1}(\beta_S(x_i)))} + \frac{(S-1)F(\beta_N^{-1}(\beta_S(x_i)))^{N}F(x_i)^{(S-2)}}{\beta_S'(x_i)}\right)(x_i - \beta_S(x_i))$$
$$= F(\beta_N^{-1}(\beta_S(x_i)))^{N}F(x_i)^{(S-1)}$$

Assuming that $\beta_N(x_i) \geq \beta_N(x_i)$, I define $x_i' = \beta_N^{-1}(\beta_S(x_i))$ so that $\beta_N(x_i') = \beta_S(x_i)$ and so that it follows $x_i' \leq x_i$. The first-order conditions of both types can be rewritten as follows:

$$\left(\frac{(N-1)F(x_i')^{(N-2)}F(x_i)^{S}}{\beta_N'(x_i')} + \frac{SF(x_i')^{(N-1)}F(x_i)^{(S-1)}}{\beta_S'(x_i)}\right)(x_i' - \beta_N(x_i'))$$
$$= F(x_i')^{(N-1)}F(x_i)^{S}(1 + \mathcal{W}_e'(\beta_N(x_i')))$$
$$\left(N\frac{F(x_i')^{(N-1)}F(x_i)^{(S-1)}}{\beta_N'(x_i')} + (S-1)\frac{F(x_i')^{N}F(x_i)^{(S-2)}}{\beta_S'(x_i)}\right)(x_i - \beta_S(x_i))$$
$$= F(x_i')^{N}F(x_i)^{(S-1)}$$

And after some further reformulations I obtain the following:

$$\begin{aligned}\left(S\frac{F(x_i')^{(N-1)}}{F(x_i)}\frac{\beta_N'(x_i')}{\beta_S'(x_i)} + (N-1)F(x_i')^{(N-2)}\right)(x_i' - \beta_N(x_i')) &= F(x_i')^{(N-1)}\beta_N'(x_i')\frac{N-1}{N}\\ \left(N\frac{F(x_i)^{(S-1)}}{F(x_i')}\frac{\beta_S'(x_i)}{\beta_N'(x_i')} + (S-1)F(x_i)^{(S-2)}\right)(x_i - \beta_S(x_i)) &= F(x_i)^{(S-1)}\beta_S'(x_i)\end{aligned} \tag{21}$$

Setting $N = S = A$ and taking a ration of these two FOCs and after some reformulations I obtain the following:

$$\frac{A-1}{A}(x_i - \beta_S(x_i)) = (x_i' - \beta_N(x_i'))$$

Which proves that $\beta_S(x_i) \neq \beta_N(x_i)$ and, for an increasing equilibrium bidding strategy, that $\beta_N(x_i) > \beta_S(x_i)$.

A comparison of the first condition in (21) to the first-order condition when N agents subject to the equal loss allocation are bidding on their own in equation (10) reveals that the agent is subject to a loss allocation bid at least as high when the group which is not subject to loss allocation is added.

Reduce the CCP's expected losses

Strategy of proof: just show for one example how losses will be lower, concentrate thereby on the β where all do not expect to pay.

Lower ex ante losses

Proof-Strategy: There are two possible outcomes. Either someone subject to the loss allocation wins or someone not subject to the loss allocation wins. Simply show for both cases that the ex-ante loss of the losers will be lower in both cases compared to where only agents subject to the loss allocation participate in the auction.

Proof of Proposition 4

I need to show that the participation constraint is always satisfied and that the termination constraint is always satisfied whenever the payment of the loser is lower than the maximal losses the CCP can inflict on the leaving agent.

Participation constraint

The expected loss when not participating in equation (6) can be calculated for all loss allocation arrangements based on equation (17) as follows:

$$\begin{aligned}
\int_{v_A-1}^{v_A} \mathcal{L}(\beta(y))dG(y) &= \int_{v_A-1}^{v_A} \Big(-\frac{1}{N+1}y - \frac{1}{N(N+1)}(v_A-1) + \frac{v_D-(\gamma+\delta+\epsilon)}{N}\Big) dG(y) \\
&= -\frac{1}{N+1}\int_{v_A-1}^{v_A} y dG(y) + \int_{v_A-1}^{v_A} \Big(-\frac{1}{N(N+1)}(v_A-1) + \frac{v_D-(\gamma+\delta+\epsilon)}{N}\Big) dG(y) \\
&= -\frac{N-1}{N+1}(\frac{v_A-1}{N-1} + \frac{1}{N}) - \frac{1}{N(N+1)}(v_A-1) + \frac{v_D-(\gamma+\delta+\epsilon)}{N} \\
&= \frac{1}{N}\big((v_D-v_A) - (\gamma+\delta+\epsilon) + \frac{2}{N+1}\big)
\end{aligned}$$

The expected profit for any loss allocation arrangement in equation (17) is replicated here again as follows:

$$\pi(x_i) = \frac{(1-c_i)^N}{N} - \frac{1}{N}\big((v_D - v_A) - (\gamma+\delta+\epsilon) + \frac{2}{N+1}\big) \geq -\int_{v_A-1}^{v_A} \mathcal{L}(\beta(y))dG(y)$$

so that it is easy to show that the inequality always holds since the expected profit for any loss allocation is at least as large as the losses that the CCP can inflict on the non-participating agent.

Termination constraint

For which bidder / loser combination is the termination constraint (7) most limiting? First, given the winning bid $\beta^*(x_i)$, then the termination constraint is most limiting where the losers have private costs very close to the winning bidder, i.e., $\lim c_{-i} \to c_i$. Second, given that $c_{-i} = c_i$, then the termination constraint is most limiting where the winner has a private cost of zero, i.e., $c_i = 0$ because

$$\mathcal{L}(\beta(x_i)) - c_i = -\frac{1}{N+1}x_i - \frac{1}{N(N+1)}(v_A - 1) + \frac{v_D - (\gamma+\delta+\epsilon)}{N} - c_i$$

is highest for $c_i = 0$.
Therefore, the TC will never be violated when I can show that $\mathcal{L}\beta(v_A) \leq l$.

Proof of Proposition 6

For $N_A < N$ participating in the auction but all N agents being subject to the loss allocation, the budget constraint of the CCP is as follows:

$$(v_D - \beta^*_{N_A}) - (\gamma+\delta+\epsilon) = (N-1)\mathcal{L}_{N_A}(\beta^*_{N_A}) + \mathcal{W}_{N_A}(\beta^*_{N_A})$$

The first-order conditions can be easily derived from equation (4), as follows:

$$\frac{N}{N-1}g_{N_A}(x_i)Z_{N_A}(x_i) + G_{N_A}(x_i)Z'_{N_A}(x_i) = g_{N_A}(x_i)\Big(x_i + \frac{v_D - (\gamma+\delta+\epsilon)}{N-1}\Big)$$

with $G_{N_A}(x_i) = (x_i - (v_A - 1))^{N_A - 1}$ and $Z_{N_A}(x_i) = \beta_{N_A}(x_i) + \mathcal{W}_{N_A}(\beta_{N_A}(x_i))$ leading to the solution

$$Z_{N_A}(x_i) = \frac{N-1}{N + \frac{N-1}{N_A-1}} x_i + \frac{N-1}{N(\frac{N(N_A-1)}{N-1} + 1)}(v_A - 1) + \frac{v_D - (\gamma + \delta + \epsilon)}{N}$$
$$\mathcal{L}_{N_A}(x_i) = -\frac{1}{N + \frac{N-1}{N_A-1}} x_i - \frac{1}{N(\frac{N(N_A-1)}{N-1} + 1)}(v_A - 1) + \frac{v_D - (\gamma + \delta + \epsilon)}{N}$$

Relating this to equation (17), it can be shown that $Z_{N_A}(x_i) < Z(x_i)$ and $\mathcal{L}_{N_A}(x_i) > \mathcal{L}(x_i)$ for $N_A < N$, so the profit of the invited agents is always larger:

$$\underbrace{\pi_{N_A}(x_i)}_{\text{expected profit of invited agents}} > \underbrace{-\int_{v_A-1}^{v_A} \mathcal{L}(\beta_{N_A}(y))dG(y),}_{\text{expected profit of non-invited agents}} \quad \text{for } x_i > v_A - 1$$

References

M. R. Baye, D. Kovenock, and C. G. de Vries. Rigging the lobbying process: An application of the all-pay auction. *The American Economic Review*, 83(1):289–294, 1993. ISSN 00028282. URL `http://www.jstor.org/stable/2117512`.

CPMI-IOSCO. Principles for financial market infrastructures. Technical report, Bank for International Settlement, 2012.

CPMI-IOSCO. Recovery of financial market infrastructures. Technical report, Bank for International Settlement, 2017.

CPMI-IOSCO. A discussion paper on central counterparty default management auctions. Technical report, Bank for International Settlement, 2019.

CPSS. A glossary of terms used in payments and settlement systems. Technical report, Bank for International Settlement, 2016.

G. Ferrara and X. Li. Central counterparty auction design. Bank of England working papers 669, Bank of England, Aug. 2017. URL `https://ideas.repec.org/p/boe/boeewp/0669.html`.

W. Huang and H. Zhu. Design of central counterparty default management auctions. Unpublished Manuscript, unpublished.

T. P. Hubbard and R. Kirkegaard. Asymmetric auctions with more than two bidders. Working Papers 1502, University of Guelph, Department of Economics and Finance, 2015. URL `https://ideas.repec.org/p/gue/guelph/2015-02.html`.

V. Krishna. *Auction Theory*. Elsevier, 2 edition, 2010. URL `https://EconPapers.repec.org/RePEc:eee:monogr:9780124262973`.

R. S. Kroszner. Can the financial markets privately regulate risk?: The development of derivatives clearinghouses and recent over-the-counter innovations. *Journal of Money, Credit and Banking*, 31(3):596–618, 1999. ISSN 00222879, 15384616. URL `http://www.jstor.org/stable/2601077`.

E. Maskin and J. Riley. Asymmetric Auctions. *The Review of Economic Studies*, 67(3):413–438, 07 2000. ISSN 0034-6527. doi: 10.1111/1467-937X.00137. URL `https://dx.doi.org/10.1111/1467-937X.00137`.

F. Umar, W. Huang, and E. Takats. Clearing risks in otc derivatives markets: the ccp-bank nexus. *BIS Quarterly Review*, 2018.

A. R. Valukas. Lehman brothers holdings inc. chapter 11 proceedings examiner report. Technical report, Jenner & Block, 2010. URL `https://jenner.com/lehman`.

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961. ISSN 00221082, 15406261. URL `http://www.jstor.org/stable/2977633`.

G. Vuillemey. Mitigating fire sales with contracts: Theory and evidence. *SSRN Electronic Journal*, 01 2019. doi: 10.2139/ssrn.3355142.

Chapter 4

Markets in the Presence of Moral Hazard, Limited Liability and Nonexclusive Contracts

Markets in the Presence of Moral Hazard, Limited Liability and Nonexclusive Contracts*

Robert Oleschak
Swiss National Bank (SNB)
Bundesplatz 1, CH-3003 Bern
robert.oleschak@snb.ch

June 19, 2020

Abstract

When contracts are nonexclusive, agents might find it profitable to trade too many contracts, exert low levels of effort and subsequently default. I show that collateral solves this moral hazard problem. However, due to collateral costs and positive externalities, trades in bilateral clearing are mostly uncollateralized. Contractual innovations (e.g. termination clause) cannot solve this moral hazard problem. I show that a CCP deals with the moral hazard problem by setting position limits but that this does do not work when CCPs compete for markets. Regulatory minimum collateral requirements are most robust and effective measure to overcome such moral hazard problems.

Keywords: Bilateral Clearing, Central Counterparty, Moral Hazard, Limited Liability, Regulation

JEL Classification: D82, D86, G28

*I would like to thank Thorsten V. Koeppl and Cyril Monnet for their insightful comments.
Disclaimer: The views, opinions, findings, and conclusions or recommendations expressed in this paper are strictly those of the author. They do not necessarily reflect the views of the Swiss National Bank (SNB). The SNB takes no responsibility for any errors or omissions in, or for the correctness of, the information contained in this paper.

1 Introduction

In derivatives markets, a distinction is often made between exchange-traded and over-the-counter (OTC) markets. Exchange-traded derivative markets are typically well regulated, trade in standardized and liquid contracts that are subsequently cleared by central counterparties (CCPs) subject to strict collateral rules. Whereas OTC derivatives markets trade in customized but illiquid contracts with varying levels of collateral to protect against default, generally subject to lower levels of regulation. In this setting, the size of OTC derivatives markets have grown rapidly in the run up to the financial crisis in 2008 and regulators have voiced concern with regards to how these risk are managed, including the low level of collateral posted to protect against counterparty risk (see for example CPSS [2007]). But it was only during the financial crisis that the full extent of the risks involved and their contagious effect was exposed.

In order to reduce systemic risks in the OTC derivatives markets, the G20 announced in 2009 that standardized OTC derivative contracts should be cleared through central counterparties (CCPs).[1] As a result, regulators across the globe started to mandate financial institutions to use CCPs for OTC markets that are deemed sufficiently standardized. At the same time, since 2016 rules are currently being phased-in requiring stricter margin requirements for non-cleared derivatives. These measures not only aim at making derivatives markets more safe but they are designed to explicitly incentivise CCP clearing (FSB-BIS-CPMI-IOSCO [2018]).

With this regulatory push to promote CCP clearing, a new distinction for derivatives markets emerges, namely between those derivatives markets cleared by CCPs and those not cleared by CCPs (bilateral clearing). From today's perspective, CCPs can be easily seen as constructions invented and promoted by regulators. However, it was the market participants that more than hundred years ago founded

[1] G20 Leaders agreed in September 2009 that: "All standardised OTC derivative contracts should be traded on exchanges or electronic trading platforms, where appropriate, and cleared through central counterparties by end-2012 at the latest. OTC derivative contracts should be reported to trade repositories. Non-centrally cleared contracts should be subject to higher capital requirements. We ask the FSB and its relevant members to assess regularly implementation and whether it is sufficient to improve transparency in the derivatives markets, mitigate systemic risk, and protect against market abuse."

the first entity resembling today's CCPs.[2] Currently, most of the CCPs that operate and clear a wide range of derivative and cash markets have mostly been founded by market participants to manage risks and improve settlement before regulatory requirements became an important driving force. That includes half a dozen CCPs that possess substantial fiancial resources and centrally clear important global financial markets[3], as well as dozens of CCPs that are important on a national or regional level.

Given that CCP and bilateral clearing will continue to exist side-by-side, this paper aims to study derivatives markets where (risk-averse) agents have the choice to trade contracts with risk neutral counterparties (banks) through a CCP (with strict collateral requirements) or clear them bilaterally where the collateral requirements can be negotiated, i.e. in a sense there is not only competition between CCPs but between CCPs and bilateral clearing. An important feature of the model is that without collateral, the agents cannot control the size of positions taken by the bank, i.e. where financial contracts are non-exclusive and that the bank's likelihood of default is endogenously defined by the size of the position.

In bilateral clearing, the agent thus faces a trade-off between charging costly collateral to manage the risk of nonperformance. In CCP clearing this task is delegated to the CCP.

I proceed by starting with a simple setting and then extend the model step-by-step. The first result in the paper is that if the agent was able to impose exclusivity in bilateral clearing, she would prefer a contract without charging any collateral, given that default-risk of the bank is low enough, the opportunity cost of collateral high or the risk-aversion of the agent sufficiently low.

However, if nonexclusivity of contracts is assumed, then I show that in big markets (i.e. in markets where banks have the opportunity to trade many contracts) the agent would need to charge high margins to avoid that the bank signs too many contracts. But charging collateral exhibits positive externalities and there might be no market. Namely, if all agents charge collateral, then a single agent can safely

[2]See for example Vuillemey [2020] who discusses the creation of the Caisse de Liquidation des Affaires en Marchandises in the market for coffee futures in Le Havre (France) in 1882.

[3]For example, Chicago Mercantile Exchange, ICE Clear, LCH, Eurex Clearing, Japan Securities Clearing Corporation, the Options Clearing Coprporation, and Fixed Income Clearing Corporation

deviate and not charge collateral at all in the knowledge that the bank cannot and will not sign too many contracts.

In an extension of the bilateral clearing model, I show that contractual innovations like termination clauses proposed for example by the International Swaps and Derivatives Association (ISDA) can extend the market size where it is safe to trade contracts without collateral, but at some point charging collateral becomes inevitable.

At this point CCPs enter the stage. Due to risk pooling they can offer a default-free contract for less collateral. But this is not the only advantage a CCP can offer. Given that all agents choose to clear through a sole CCP, the CCP is in a position to set trading limits for banks and thus limit the moral hazard problem of signing too many contracts too. CCP clearing, however, does not always form an equilibrium for two reasons. The first reason is competition between CCPs. If more then one CCP is active in a market, then banks can circumvent the limits that CCPs set and exhibit higher rates of default. The second reason are deviating agents. If the opportunity costs of collateral is too high, agents might still prefer to clear bilaterally without collateral and deviate if all other agents clear with a CCP.

CCPs have tried to solve this issue by implicitly imposing mandatory clearing for certain liquid (and thus big) markets and allowing only one CCP to do the clearing, e.g. it is not possible to trade on a derivative market without using the dedicated CCP which in many cases is an integral part of the exchange market. However, this strategy might not work for contracts that are traded at many marketplaces globally where it is difficult to impose de-facto mandatory clearing or where competition between CCPs in different jurisdictions allow banks to distribute their contracts.

Another situation where self-organized mandatory clearing with one CCP is difficult to impose is in markets which have grown from small to a big in a relatively short time frame. In such cases the market participants might not react quickly enough and establish market practices or even set up a CCP taking into account the growing importance and market size. This has arguably happened in the market for credit default swaps (CDS) which has rapidly grown in a few years and CCPs were set up only after AIG reported massive losses on CDS positions that

lead to the bailout of the company (see for example Arora et al. [2012] for a discussion of a discussion of the CDS market). In contrast, LCH launched in 1999 a CCP for interest rate swaps (IRS) with wide support from global banks, covering by the end of 2006 around 40% of the global inter-dealer IRS market (CPSS [2007]).
Finally, I show that collateral discourages banks from trading too many contracts, encourages the bank to exert high levels of effort, and ensures that the obligations are always met. But collateral exhibits positive externalities and is costly, therefore even risk-averse agents would prefer not to use it. I show that contractual innovations (e.g. termination clauses) do not mimic the function of collateral satisfactorily and cannot solve the moral hazard issue. Regulation requiring agents to exchange minimum levels of collateral helps to overcome this problem.

Related Literature

The analysis of competition between multiple exchanges has brought forward several modeling frameworks. Santos and Scheinkman [2001a] consider whether competition among exchanges lead to excessively low standards. They show in a framework where exchanges design contracts and which allows for the default of the trader but not for them to trade in multiple markets that the use of collateral can lead to a constrained efficient outcome. However, in Santos and Scheinkman [2001b] they note, that under nonexclusivity the equilibrium might not hold any more. Bisin and Guaitoli [2004] consider competitive equilibria where agents trade in standardized markets non-exclusive contracts under a wide set of asymmetric information. They show that two prices (a bid and ask price) and a 'pooling' asset are enough for a competitive equilibrium to exist, i.e. when markets are sufficiently complete. The pooling asset is a security whose payoff is the average total net amount due to agents who traded the contract. In this paper, I do not assume that markets are sufficiently complete but allow agents to trade incomplete contracts. This setup gives rise to innovating CCPs who not only can limit the amount of trades that banks execute but offer default-free contracts.
The study of non-exclusive contracts is usually combined with some form of moral hazard behaviour. For example Bizer and DeMarzo [1992] study environments where agents borrow sequentially from more than one lender, and so the borrower

has an incentive to approach more than one banks for further borrowing. This non-exclusivity is discussed by Bisin and Guaitoli [2004] study equilibria for economies characterized by moral hazard (hidden action) and non-exclusive contracts. This paper borrowed several modelling aspects from these two papers. Unlike in these two papers, where the risk-averse agent is subject to moral hazard, in my paper it is the risk-neutral agent who is subject to moral hazard.

More generally, the literature on central clearing has grown rapidly following the financial crisis of 2008. One branch of study focuses on the systemic aspect of CCP clearing by comparing counterparty risks in an environment where financial contracts are netted bilaterally between financial institutions (bilateral clearing) with the situation where CCPs allow for multilateral netting of traded financial contracts (central clearing). For example, Duffie and Zhu [2011], Cont and Kokholm [2014], and Lewandowska [2015] analyse whether central clearing can reduce counterparty risks compared to bilateral clearing. Another branch of papers focuses on the disincentives present in markets cleared by CCPs. Huang [2019], for example, assesses the appropriate size of collateral and equity from the viewpoint of a profit-oriented CCP subject to limited liability or Bignon and Vuillemey [2017] show, based on historical empirical evidence, that CCPs might have in the past delayed declaring an agent into default and in hopes of a resurrection. Other papers (for example Angelo et al. [2019] or Arora et al. [2012]) analyse the effects of mandating CCP clearing on interest rates or credit markets. In this paper, following Biais et al. [2016] I model the CCP as an entity that pools risks and sets collateral to guarantee default-free contracts. The trading game in this paper is inspired by Leitner [2012] who proposes a central mechanism that sets position limits and reveals the names of agents who hit these limits and so disciplines agents not to sign too many contracts. In my paper I go one step futher and consider environments where there is more than one CCP.

2 The Basic Model

There are two periods. The economy is populated by a continuum of risk-averse agents and risk-neutral banks. The agents are endowed with one unit of an identical, illiquid, risky asset yielding a publicly observable stochastic return R at $t = 1$.

The banks are endowed with one unit of wealth which they can invest at $t = 0$ into a risky asset paying a privately observable stochastic return θ at $t = 1$ or into a safe asset (cash) yielding zero return. If the bank invests into the risky asset she can affect her return privately by choosing high or low effort $e \in \{a, b\}$. Agents can enter at $t = 0$ nonexclusive state-contingent contracts with banks for insurance purpose, i.e. agents cannot control with how many other agents a bank signs a contract. Banks are protected by limited liability.

2.1 Preferences and Technology

The utility of the agents is given by function u, where $u'' < 0$. The agents cannot influence the return of their assets which depend on the state of the economy and can take two values $R_H > R_L$ with $\Delta R = R_H - R_L = 1$, and where $p = \frac{1}{2}$ is the probability of high return. The outcome of R is public information. R_H can be interpreted as the outcome of a good state of the economy and R_L as the bad state. The expected utility can be expressed as

$$U(R) = \frac{1}{2}u(R_H) + \frac{1}{2}u(R_L)$$

Each bank can invest her one unit of wealth in period 0 either into a riskless asset with zero net return (i.e. cash) or into a risky asset with random return θ that is i.i.d. across banks, whose realization is only privately observable, and takes two values $1.5 > \theta_H > 1$ and $\theta_L = 0$.[4] The bank can take action e that affects the probability distribution of the investment into the risky asset. The action taken at $t = 0$ can take two values $e \in \{a, b\}$, is only privately observable, and taking action e carries dis-utility $C_e = (e - b)$ for the bank. For simplicity, I assume that the action equals the probability of the good outcome, and that action a is the high effort action, where $1 > a > b > 0.5$. The expected utility of a bank which has invested the share $\alpha \in [0, 1]$ of her wealth into the riskless asset and a share

[4] I allow the high return θ_H to be within a certain range only. This limits the number of possible solutions without affecting the main results of the paper. Assuming that a return is lower than 50% is in any case not a very limiting assumption.

$(1-\alpha)$ of her wealth into the risky asset and taken action e is:

$$\pi(\alpha, e) = \alpha + (1-\alpha)(\theta_H e - C_e)$$

Note, that the asset under management and the dis-utility are connected: if the bank invests only a share $(1-\alpha)$ of her wealth into the risky asset then the dis-utility only applies to this part of the investment (i.e. keeping assets in cash does not cost any effort). C_e enters as dis-utility for the bank only: it does not diminish the capital that can be invested, nor the resources available to transfers to the agents.

The values have been set so as to simplify often used expressions and the final results: First, the expected return in case of low effort is defined to be $\theta_H b = 1$. Second, I define $r_e = \theta_H e - 1 - C_e$ where r_e represents the expected net return on the risky asset after discounting for the dis-utility connected to action e. Based on these specifications $r_b = 0$ and $r_a = \frac{(\theta_H a - 1)(\theta_H - 1)}{\theta_H} > 0$.

2.2 Market for State-Contingent Contracts

The agent knows the banks technology consisting of choice e but cannot observe the level of effort. The number of contracts a bank has signed or will sign in future cannot be observed by the agent either. The agent may enter into a *standard swap contract* with one bank at $t = 0$. The contract consists of $\langle F, \alpha \rangle$ specifying the swap rate F and the collateral α that the bank must place in an escrow account and cannot use for investment purposes.[5] The contract implies that the agent expects to receive a transfer $F - R_H < 0$ in the good state of the economy or $F - R_L > 0$ in the bad state of the economy, thus a positive transfer indicates a transfer from the bank to the agent and vice-versa.

I allow the collateral to take two values $\{0, F - R_L\}$, i.e. the agent can either decide not to collateralize the trade or to fully collateralize it.[6] The collateral can

[5] The contract will not be contingent on the return θ of the bank. This is because it is difficult or impossible for the agent to monitor the return of the bank.

[6] Note, that I restrict the contract on two accounts. First, I allow only a standard swap contract to be written, instead of allowing the agent to fix different transfers for all four possible states of the economy and second, I allow the collateral to take two values only. Both limitations simplify the results but do not affect the general results presented in this paper.

be accessed by the agent only if the bank does not fulfill its contractual obligations. If all contractual obligations are fulfilled, the collateral will be returned to the bank in period 1. The bank is protected by limited liability and therefore the maximum amount she can transfer at any state of the world to the agent is limited by what is tangible. I assume that the bank does not suffer any penalty (other than losing her collateral placed in the escrow account) if she defaults on her contractual obligations.

2.3 The Trading Game

The following trading game is based on Leitner [2012] but includes some additional elements. At $t = 0$ there are N trading rounds. In the first trading round a fraction $\frac{1}{N}$ of agents and of banks are chosen randomly to enter the market. The agent proposes a state-contingent contract and the bank can accept or reject. If the agent's contract has been accepted she will exit the market or else stay for the next round.[7] However, agents do not know in which round they are in.

A bank however, can stay in the market for the next round even if she accepted a contract. In the next round a new bank can only enter if one bank from the preceding period has left. This means that the banks which have randomly been chosen in the first round can stay in the market for up to N trading rounds and sign N contracts with N different agents. N can be interpreted as the size of the market.

Once a bank has accepted $n \leq N$ contracts and left the market she invests all remaining wealth $(1 - n\alpha)$ into the risky asset θ and chooses the effort e she wants to apply.[8]

2.4 Definition of an Equilibrium

The outcome of the game is defined by $\langle n, F, \alpha, e \rangle$. Given such an outcome in period 1, the utility to the bank that has accepted n contracts and invested is

[7]In equilibrium, agents will propose a contract which the bank will always accept and thus agents will stay only for one round.

[8]I will not allow the bank to consume wealth $(1 - n\alpha)$ in period 0. The bank can only invest into the risky or riskless asset.

$\pi(n, F, \alpha, e) = E[(n\alpha + (1 - n\alpha)\theta e - n(F - R))^+] - (1 - n\alpha)C_e$. In contrast, banks that have signed no contract have an expected utility of $\pi(0, 0, a) = \theta_H a - C_a$. The agent's utility can be expressed as $U(R, F) = E[u(R + \min(F - R, \alpha + \frac{(1-n\alpha)\theta}{n})]$ consisting of the return of the risky asset and the contractual transfer, respectively if the bank is unable to fulfill her contractual agreements, then the agents receive the collateral and share the proceeds from the investment. Thus, the actual transfer consists of either the agreed transfer $F - R$ or, if the bank has not sufficient resources, the agent receives the collateral and a share of any remaining resources.

A strategy for a bank is a policy defining which set of transfer contracts $\langle F, \alpha \rangle$ she is willing to accept for how many rounds n, together with associated investment and effort choices e. A strategy for an agent is a set of transfer contracts $\langle F, \alpha \rangle$ she is willing to propose.

In equilibrium, the policy of each bank should maximize profits given the strategy of the agent, and the strategy of the agent should maximize utility given the banks' policies. I will apply the concept of subgame-perfect equilibria in order to rule out non-credible strategies. In any subgame-perfect equilibrium, a bank accepts only transfer contracts that earn profits at least as good as not accepting a contract. Profits on a transfer contract depend on the effort of the bank, which is not known to the agent. From a subgame perfection, the bank always chooses effort optimally given the transfer contract.

To better organize and present the results, I will first discuss the well-known case where contractual relationship are exclusive and one bank contracts with one agent only ($N = 1$). I will then move to the case where the agent cannot condition the transfer contract to the bank's other contractual obligations, i.e. where contractual relationships are nonexclusive and the bank can negotiate contracts bilaterally with several agents. Further, I will analyze the situation, where a central counterparty interposes itself between agent and bank and guarantees the performance of the contract. Finally, I allow bilateral and CCP clearing to exist side-by-side and analyze the implications for trading equilibria.

3 Exclusive Contracts

The contractual relationship between an agent and a bank is exclusive in the sense that the bank as well as the agent can write only one contract. I will first formally describe the problem solved by the bank and the agent, define of the equilibrium, discuss it and relate it to derivative markets.

3.1 Maximization problem of banks and agents

Given that the bank accepts the contract $\langle F, \alpha \rangle$ proposed by the agent she chooses the optimal action $e \in \{a, b\}$ to maximize

$$\pi(F, \alpha, e) = E[\big(\alpha + (1-\alpha)\theta e - (F - R)\big)^+] - C_e \tag{1}$$

Equation (1) contains two elements: limited liability and moral hazard which I will discuss each in turn.
Limited liability requires that the transfers $\tau_H = F - R_H$ and $\tau_L = F - R_L$ satisfy the following four inequalities:

$$\begin{array}{rll} (R_H, \theta_H): & \alpha + (1-\alpha)\theta_H & -\tau_H \geq 0 \\ (R_H, \theta_L): & \alpha & -\tau_H \geq 0 \\ (R_L, \theta_H): & \alpha + (1-\alpha)\theta_H & -\tau_L \geq 0 \\ (R_L, \theta_L): & \alpha & -\tau_L \geq 0 \end{array}$$

Since the agent prefers to receive payments if her return is low R_L (i.e. $\tau_L > 0$) and transfer payments to the bank in case of high return R_H (i.e. $\tau_H < 0$) limited liability in the first two cases does not bind. In the third case, limited liability can bind, if return θ_H is too small compared to the required transfer τ_L or if the share of collateral α is too high, so that the bank does not have sufficient resources to invest into high-yielding projects. In the fourth case limited liability always binds.
Moral hazard with transfer contract occurs when the bank prefers to exhibit low effort b instead of high effort a after signing a transfer contract. A necessary condition to avoid moral hazard (i.e. to ensure that the bank prefers to exhibit

high effort $e = a$ even after signing transfer contract) is:

$$\frac{1}{2}(\tau_L - \alpha) \leq (1-\alpha)(\theta_H - 1) \tag{2}$$

The incentive constraint binds before the resource constraint in case of (R_L, θ_H) binds.[9] That means whenever the bank finds it optimal to exhibit high effort $e = a$ the resource constraint in case of (R_L, θ_H) does not bind. The incentive constraint never binds in case of full collateralization because the left-hand side is by definition zero. This is an important statement since it means that in case of full collateralization the agent not only ensures that she will always be paid but that it is always in the interest of the bank to exert high effort.
The agent proposes contract $\langle F, \alpha \rangle$ to maximize

$$U(R, F) \tag{3}$$

subject to

$$\pi(F, \alpha, e) \geq \pi(0, 0, a) \tag{4}$$

$$-R \leq \tau_R \leq \alpha + (1-\alpha)\theta, \quad \forall\ R, \theta \tag{5}$$

$$e \text{ solve } (1) \tag{6}$$

The *participation constraint* in equation (4) requires that the bank accepting the contract is at least as well off with the contract compared to not having the contract at all. The *resource constraint* in equation (5) incorporates the limited liability of the bank (rhs) as well as the limited resources that the agent can transfer to the bank (lhs). The lhs of the equation never binds, i.e. the agent will never propose a contract, where she will transfer to the bank all her returns. As to the rhs of the equation: In the exclusive world, the agent will need to resolve the trade-off between charging costly collateral α and accepting lower (or no) transfer in case both the investment of the bank and of the agent turn bad (i.e. R_L, θ_L). Finally, (6) accounts for the interdependency between the agent's and the bank's

[9] To see this, compare the incentive constraint $\frac{1}{2}(\tau_L - \alpha) \leq (1-\alpha)(\theta_H - 1)$ with the resource constraint $\alpha + (1-\alpha)\theta_H - \tau_L \geq 0$ in case of (R_L, θ_H). After some reformulations I can show that $2(\theta_H - 1) > \theta_H$ (because $\theta_H < 1.5$) which proves that the IC binds first.

maximization problem: A contract, if accepted, can affect the level of effort of the bank, which in turn has an effect on the utility of the agent. I will refer to it as the *incentive constraint*.

3.2 Equilibrium: Definition and Results

The following definition of the equilibrium is standard and has already been used for example by Bisin and Guaitoli [2004] and others:

Definition 1 *An exclusive contract equilibrium is an array $\langle F, \alpha, e \rangle$ such that e maximize (1) given $\langle F, \alpha \rangle$ and $\langle F, \alpha \rangle$ maximize (3) subject to (4)-(6).*

I will now discuss the solution to the incentive unconstrained problem. The following results and the analysis are graphically illustrated.

Corollary 1 *Given action $e = a$ by the bank, the optimal contract $\langle F^*, \alpha^* \rangle$ can be expressed as*

$$F^* = \begin{cases} F_1^* = \frac{1}{(1+a)} + R_L & \text{if} \quad \alpha_1^* = 0 \\ F_2^* = \frac{1}{2(1+r_a)} + R_L & \text{if} \quad \alpha_2^* = F_2^* - R_L \end{cases} \tag{7}$$

and where the optimal level of collateral is $\alpha_1^* = 0$ if

$$\frac{1}{2}(1+a)u(F_1^*) + \frac{1}{2}(1-a)u(R_L) \geq u(F_2^*) \tag{8}$$

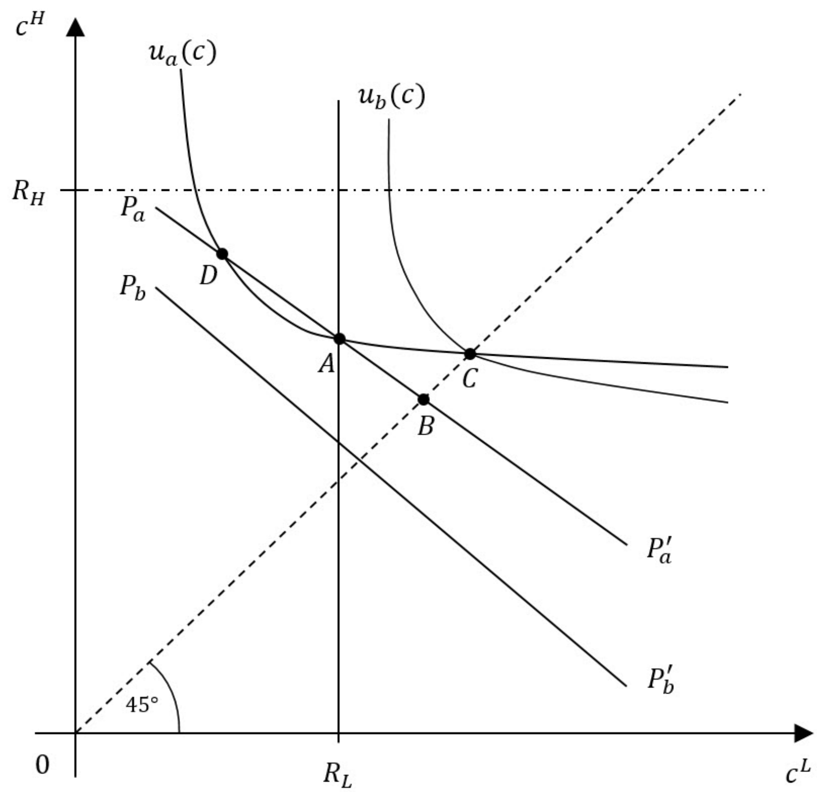
or else $\alpha_2^* = F_2^* - R_L$.

Corollary 1 captures the structure that the agent wants to impose on the risk transfer contract given that the bank chooses high action a (i.e. without considering the incentive constraint in equation (6)).

The implication of the first line in equation (7) is that the agent wants to achieve the same level of consumption $c_1^H = F_1^*$ in three out of four possible states of the world. In the fourth state of the world, where both the bank and the agent are faced with low return (i.e. R_L, θ_L) the consumption consists only of the low return $c_1^L = R_L$, because the agent has chosen no collateralization. In the second line, the agent chooses full collateralization. In this case her consumption is the same

in all four different states of the world: $c_2^H = c_2^L = F_2^*$.
The interplay between this particular structure of the risk-transfer contract and the participation, resource, and incentive constraints expressed in equations (4) to (6) is depicted in Figure 1. The x-axis and y-axis depict the consumption in all four different states of the world: To be able to depict all states of the world in a two-dimensional graph, I combine on the y-axis the three states of the world, where the agents reaches constant consumption after signing the contract $c_1^H = F_1^*$ or $c_2^H = F_2^*$ and on the x-axis I depict the fourth state of the world, where the bank would need to pay, but has zero return and therefore only the collateral is available for transfer, i.e. $c_1^L = R_L$ or $c_2^L = R_L + \alpha_2^*$.
The two lines P_a, P_a', and P_b, P_b' represent the participation constraint of the bank conditional on action $e \in \{a, b\}$ being chosen by the bank. The indifference curves of the agents u_a and u_b have the same expected utility to make them comparable. Point A depicts the level of utility, that the agent can achieve with the transfer contract $\langle F_1^*, \alpha_1^* = 0\rangle$ (no collateralization). In fact, the agent could achieve a higher utility, if she chose a point between A and D but since she cannot charge negative margin, $\alpha_a = 0$ is the best she can achieve. Point B can be reached with the full collateralization contract $\langle F_2^*, \alpha_2^* = F_2^* - R_L\rangle$ but is in this example inferior compared to no collateralization.
Note, that the bank is willing to choose action a, since choosing action b would lead to a lower expected return for the bank (since the participation constraint P_bP_b' lies below point A). Note also, that the two indifference curves intersect exactly at the 45 degree line, because in this case the agent is indifferent whether the bank chooses high or low effort. As can be shown, it is never optimal for the agent to propose a contract which lies below the 45 degree line (i.e. since it is never optimal for the agent to charge more collateral to make her better off in the fourth state of the world). In addition, any consumption without collateral must lie at the vertical line going through R_L and any consumption with full collateral lies on the 45 degree line. If continuous collateral was allowed, then solutions between those two lines would be possible too. Finally, the agent would never suggest a contract where $F - R_H > 0$ (i.e. receiving a positive transfer in case of a good state of the economy), therefore a contract must lie always below the horizontal line intersecting R_H.

Figure 1: Utility of agents and participation constraint as well as level of effort by banks



Whether the agent chooses no or full collateralization depends on whether equation (8) holds or not. I argue, that $\alpha_1^* = 0$ is an optimal outcome for a wide range of possible parameters. Consider the agent's utility $u(x) = -e^{-cx}$, where c is the constant absolute risk aversion. With a minor refomulation equation (8) then can be expressed as:

$$\frac{1}{2}(1+a)e^{-\frac{c}{1+a}} + \frac{1}{2}(1-a) \leq e^{-\frac{c}{2(1+r_a)}}$$

It can be shown, that the inequality holds, i.e. that no collateral will be called when a is high (high likelihood that the bank will not default), with high θ_H (when calling collateral is very costly due to high opportunity costs of the bank), and low c (low risk aversion of the agent). For example, when $a = .98$ and $\theta_H = 1.3$ then with only a very high risk aversion of $c \geq 6.45$ a contract with full collateralization

is preferred.
This discussion leads to the following proposition.

Proposition 1 *In an exclusive contract environment there can be an equilibrium only with high level of effort $e^* = a$. As long as a is close to 1, θ_H high enough or risk aversion is not too high,* ***and*** *the incentive constraint $2(\theta_H - 1) \geq \frac{1}{1+a}$ is satisfied then $\langle F_1^*, \alpha_1^* = 0\rangle$ (no collateralization) is the preferred contract.*
When the incentive constraint would be violated without collateral, then agents either suggest a contract with full collateralization $\langle F_2^*, \alpha_2^* = F_2^* - R_L\rangle$ or prefers no contract at all. Finally, low effort $e = b$ can never be an equilibrium in an exclusive contract environment.

The incentive constraint $2(\theta_H - 1) \geq \frac{1}{1+a}$ in case of a contract with no collateral can be derived from combining equation (2) and the first line of equation (7). When an agent prefers a fully collateralized contract, then the bank always has an incentive to exert high effort because she earns a lower return on the invested asset but does always have to pay the transfer because it is fully collateralized.
If an agent would prefer to have a contract with zero collateralization (i.e. $U(F_2^*, R) \leq U(F_1^*, R)$) but needs to fully collateralize to ensure that the bank exerts high effort, the question is whether she would rather choose to have no contract at all., i.e. when is the following equation satisfied

$$U(R) \geq U(F_2^*, R)$$

Above inequality is satisfied when the agent has a very low risk aversion or the opportunity cost of charging collateral is high. But then the agent might prefer not to trade a contract at all, rather than charging full collateral.
Finally, the result that $e = b$ can never be an equilibrium is intuitive: If the bank exerts low effort, than $r_b = 0$ which means that charging collateral carries no opportunity costs. In this case, the agent can as well fully collateralize the trade. In such a case, the bank has an incentive to exert high effort as shown above. Low effort can never be an equilibrium.
As stated in the following corollary, an agent always prefers to have a contract with no collateralization compared to no contract at all.

Corollary 2 *Given the optimal contract $\langle F_1^*, \alpha_1^* \rangle$ the agent is always strictly better off with the contract as long as $a > 0$, i.e.*

$$U(R, F_1^*) > U(R) \text{ for } \forall \ a \in [0, 1) \tag{9}$$

From corollary 2 it follows that if the agent is sufficiently risk averse and rather prefers full collateralization that such an agent strictly prefers such a contract than no contract at all.

Cenedese Gino and Vasios [2018] study derivatives prices of interest rate swaps traded in OTC markets and find, that the swap rate decreases by posting margins and with higher buyer's creditworthiness. Both results are consistent with the findings presented in this section. In addition, I provide a theory, why agents would not ask for collateral at all, something that was empirically found in the paper.

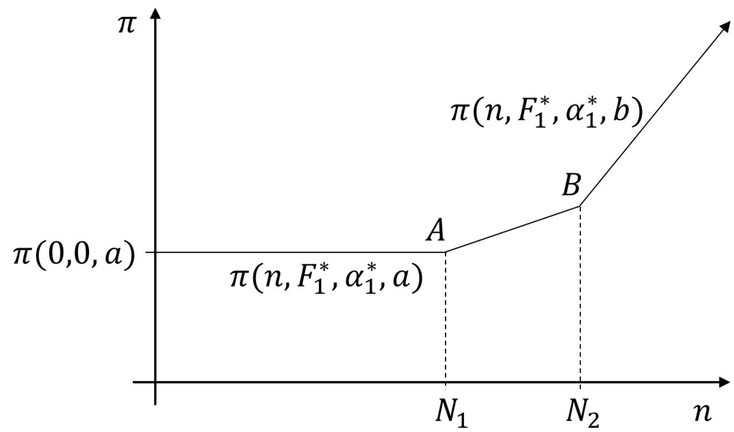
4 Nonexclusive Contracts

In this section, the contractual relationship between agent and bank is not considered to be exclusive any more: banks can have contractual relationships with more than one agent. In order to illustrate the issue assume for a moment that the agent naively observes a bank's capital, knows the banks choices regarding its effort level and associated costs and works out that it can propose the contract $\langle F_1^*, \alpha_1^* \rangle$ with no collateralization since the agent expects that the bank exerts high-level effort $e^* = a$ in case of a contract with only one counterparty.

However, suppose that if the bank can sign up to N contracts. Then the incentive constraint from proposition 1 can be expressed as $N\frac{1}{1+a} \leq 2(\theta_H - 1)$. Thus if the bank signs $N > N_1 = \lfloor 2(1+a)(\theta_H - 1) \rfloor$[10] contracts then the incentive constraint is not met any more in which case the bank can increase her profit by exerting effort $e = b$ after point A in Figure 2. In addition, if the bank signs more contracts $N > N_2 = \lfloor (1+a)\theta_H \rfloor$ she eventually violates the resource constraint in case of R_L, θ_H and her profit will increase even steeper with each additional contract after point B.

[10]The floor function $\lfloor x \rfloor$ gives the largest integer less than or equal to x.

Figure 2: Expected profit of bank with n contracts



Interpreting the number of trading rounds N as the size of the market for contingent-contracts I can state the following definition.

Definition 2 *An nonexclusive contract equilibrium for market size N consists of a contract $\langle F, \alpha\rangle$ where no agent has an incentive to deviate and offer a different contract $\langle F, \alpha\rangle'$.*

Given market size N in equilibrium all agent need to offer the same contract and the bank accepting up to N contracts must still be exerting the level of effort that the contract assumes the bank would. Based on this, I can state the following results.

Proposition 2 *In case of nonexclusive contracts, two market categories can be distinguished. In a* ***small market*** *where $N \leq N_1 = \lfloor 2(1+a)(\theta_H - 1)\rfloor$ agents are free to propose contract $\langle F_1^*, \alpha_1^*\rangle$ without or $\langle F_2^*, \alpha_2^*\rangle$ with full collateralization. In a* ***big market*** *where $N > N_1$ full collateralization $\langle F_2^*, \alpha_2^*\rangle$ is an equilibrium only when $U(R, F_1^*) \leq U(R, F_2^*)$ or else there is no equilibrium.*

The function of collateral in proposition 2 is different compared to proposition 1: Agents require collateral if the market is too big in order to avoid that the counterparty signs too many contracts (or to verify that the counterparty has not already signed too many contract). The agents do so, even if they in an exclusive contract environment, would not require collateral at all.

An issue arises if the agents would prefer not to charge collateral (i.e. $U(R, F_1^*) \leq U(R, F_2^*)$) but should in order to discipline the bank (because $N > N_1$). In this case an agent has an incentive to deviate and not charge collateral if all other agents charge collateral. In such a case there can be no equilibrium.

The question is, whether (costly) collateral is the only way of avoiding banks signing too many contracts, i.e. whether agents could not offer any other contract that would have the bank reveal N. The short answer is no: agents always have an incentive to miss-report the number of counterparties they have singed contracts or plan to sign contracts with. The reason is that agents find out whether a bank has singed too many contracts only when it is too late and the bank is in default and therefore no punishment or incentive scheme is feasible.

The issue becomes more complicated, if one is willing to relax the assumption that the number of signed contracts are private information to the bank only. In the following chapter I discuss one such framework, where a rating agency can determine with some level of probability, whether the bank has singed too many contracts or not and send a good or bad signal.[11]

5 Termination Clause

5.1 Extension of the trading game

The trading game is the same as described in chapter 2. In addition, at the end of N trading rounds and after the bank has exhibited effort $e = \{a, b\}$ a rating agency evaluates each bank and sends a signal $s \in \{s_g, s_b\}$. The rating agency sends a good signal s_g if he thinks that the number of counterparties the bank is contracting with is below a critical threshold n^* and a bad signal s_b if he thinks it

[11] Another possible framework which the author has studied but chosen not to present in this paper is to allow the agent to reliably verify the number of contracts a bank has singedg at the end of period 1. The agent can punish a bank if it has signed too many contracts when the state of the economy is good by reducing the transfer amount it was supposed to pay to the bank (or even impose a fine on the bank). At the end, this framework suffers from the same shortcomings as the first one: It extends the size of the market where no collateral must be charged, but at some market size N full collateralization cannot be avoided. Interestingly, the author is not aware of any contractual arrangement that contains such a clause. The reason might be, that it is not possible or simple to determine whether a counterparty has signed too many contracts in a good state of the economy.

is above. Conditional probability of a signal is:

$$\lambda = \text{prob}[s_g \mid n \leq n^*] = \text{prob}[s_b \mid n > n^*]$$

Given that a bank has received a bad signal s_b agents that have signed a contract with a termination clause can terminate the contract with the bank and in a final round propose a contract to another bank that has received a good signal s_g.

5.2 Maximization problem and results

The critical level of contracts is defined as $n^* = N_1$ i.e. where the incentive constraint is just met. Given that the agent can credibly threat to terminate the contract in case of a bad signal s_b but will continue with the contract in case the signal is good s_g and given that the agent proposes contract $\langle F_1^*, \alpha_1^* \rangle$ a bank will keep it's number of counterparties at or below the threshold n^* as long as

$$\underbrace{\lambda[\pi(n^*, F_1^*, \alpha_1^*, a)]}_{s_g} + \underbrace{(1-\lambda)[\pi(0,0,a)]}_{s_b} \geq \underbrace{\lambda}_{s_b} + \underbrace{(1-\lambda)[\pi(N, F_1^*, \alpha_1^*, b)]]}_{s_g}$$

In the first and second term the bank keeps the number of contracts at the critical level n^* which means that she exerts a high effort and is able to pay all transfers (except in case R_L, θ_L where she defaults). In the first term the rating agency sends a good signal and the agents do not terminate the contracts. In the second term the rating agency falsely sends a bad signal and all agents terminate the contract. Since the bank exerted high effort and invested all capital her profits are not affected by the termination.[12]

In the third and fourth term the bank contracts with the maximum possible number N of counterparties, exerts low effort b and is not able to fulfill her contractual obligations in case of (R_L, θ_L) as well as (R_L, θ_H). In the third term, the rating agency uncovers this and sends a bad signal s_b so that all agents terminate the contract and the bank is left with one unit of capital. In the fourth and last term the rating agency falsely sends a good signal s_g so that the bank can make a pos-

[12]However note, that the early termination clause is only for free when $\alpha_1^* = 0$. If that is not the case, then "false" early termination carries opportunity costs $\alpha_1^* r_a$ which would need to be compensated by the principle from the beginning.

itive profit $\pi(N, F_1^*, \alpha_1^*, b) = \frac{1}{2}(1 - N(F_1^* - R_H))$.[13]
The equation can be simplified by noting that the left-hand side is the expected profit of the bank and given a binding participation constraint equals $\pi(0, 0, a) = 1 + r_a$ I get $-\frac{1}{2}N(F_1^* - R_H) \leq \left(\frac{r_a}{(1-\lambda)} + \frac{1}{2}\right)$. Entering equation (7) for $F_1^* - R_H$ the termination clause ensures that the bank keeps the number of contracts at n^* as long as the market size is at or below the following threshold.

$$N_T = \lfloor \frac{2(1+a)}{a} \left(\frac{r_a}{(1-\lambda)} + \frac{1}{2} \right) \rfloor \tag{10}$$

Clearly, the maximal market size crucially depends on λ which stands for the accuracy of the rating agency. If λ is high, then the maximal market size where agents can trade first-best contracts can increase manifolds compared to the situation where no termination clause is present in the state contingent contract.
Is the threat from the agent to terminate the contract credible? Consider this: if the inequality in equation (10) holds than no bank will overextend and sign too many contracts. In such a case, if a rating agency sends a bad signal, then agents should assume that it is a false signal and ignore it (i.e. not terminate the contract), so the threat is not credible. However, if all agents act in this way, then the bank can safely overextend since no agent will terminate in case of a bad signal. The only way out is if the agent has a free outside option i.e. the ability to sign a new similar contract with another bank who received a good signal. Then it is a weakly dominant strategy to terminate a contract in case of a bad signal, even when all banks have an incentive to behave good.
Finally, a termination clause is not useful (or will not be exercised) in case of full collateralization.

Proposition 3 *A termination clause extends the market size where the contract*

[13] The expected profit is based on the profit multiplied with the probability. I can differentiate the following four cases where in the last two cases the profit is zero. Note too, that $b\theta_H = 1$.

bp	$\theta_H - N(F_1^* - R_H) \geq 0$
$(1-b)p$	$-N(F_1^* - R_H) \geq 0$
$b(1-p)$	0
$(1-b)(1-p)$	0

$\langle F_1^*, \alpha_1^* \rangle$ *without collateral is incentive compatible to* $N_T = \lfloor \frac{2(1+a)}{a} \left(\frac{r_a}{(1-\lambda)} + \frac{1}{2} \right) \rfloor \geq N_1$.

The last observation that $N_T \geq N_1$ can be verified by considering the ratio: $\frac{N_T}{N_1} = \frac{\lfloor \frac{r_a}{1-\lambda} + \frac{1}{2} \rfloor}{\lfloor a(\theta_H - 1) \rfloor} \geq 1$. The inequality holds even when $1 - \lambda = .5$, i.e. when the rating agency has a very bad track record of picking banks with too many contracts.[14]

It is worth noting, that parties who enter into OTC derivatives very often use the ISDA Master Agreement (or some adapted domestic version) which is a document published by the International Swaps and Derivatives Association (ISDA) to provide certain legal and credit protection. Besides offering the benefit of standardization of contractual arrangement, the ISDA Master Agreement together with its annexes defines the provision of collateral and the early termination of the contract.[15] As has been shown here, such termination clauses can have a disciplining effect on the counterparties under certain circumstances.

6 Central Counterparty Clearing

In the CCP clearing economy I will first consider the case where there is only one CCP that clears all trades, i.e. the contracts are still non-exclusive but there is an agency (the CCP) which keeps track of the contracts signed by a single bank. In the next chapter I will extend the analysis by allowing agents and banks to clear at the CCP or bilaterally and by allowing two CCPs to be active in the market.

The trading game in the CCP clearing economy is the same as described in chapter 2 with a few adjustments. When submitting trades for CCP clearing, the agents can propose the swap-rate F only. Because it is the CCP that defines and collects from the bank the collateral $\alpha_{\bar{e}}^{CCP}$ which must be sufficient to cover any potential shortfall if the bank or any other bank defaults given the economy-wide level of effort $\bar{e}$.[16] In addition, the CCP limits the number of contracts n that it allows

[14] To see this the ratio can be simplified for $1 - \lambda = .5$ to $2r_a + .5 \leq a(\theta_H - 1)$. After some reformulations I get $(\theta_a - 1.5)(\theta_H - 1) + .5 \geq 0$ which is always satisfied for $a > 0$.

[15] See https://www.isda.org for an overview of the ISDA Master Agreement and Maizar et al. [2010] for the description of the Credit Support Annex to define the provision of collateral.

[16] $\alpha_{\bar{e}}^{CCP}$ has characteristics of a default fund contribution (mutualization of risks) as well as initial margin (calculated for each trade).

each bank to submit for clearing. Finally, the bank decides based on that which level of effort e she wants to exhibit. In equilibrium e must equal $\bar{e}$.
In the following I will consider the equilibrium with and without termination clause applied by the agents.

6.1 No termination clause: Maximization problem and equilibrium

Given an economy-wide default-rate of $1-\bar{e}$ the CCP knows that a share of $\bar{e}$ banks will have a return θ_H and the other share $1-\bar{e}$ will have a return of $\theta_L = 0$ and default on their obligations in case of a bad state of the economy R_L. The collateral - or as I will call it from now on - the default fund contribution of each bank is thus by definition $\alpha_{\bar{e}}^{CCP} = (1-\bar{e})\tau_L$. Given that each bank signs n transfer contracts τ_R and contributes $\alpha_{\bar{e}}^{CCP}$ to the default fund for each contract, the bank chooses the optimal action $e \in \{a, b\}$ to maximize

$$\pi(n, F, \alpha_{\bar{e}}^{CCP}, e) = \frac{1}{2} n\alpha_{\bar{e}}^{CCP} + (1 - n\alpha_{\bar{e}}^{CCP})(\theta_H e - C_e) - n[p\tau_H + \frac{1}{2} e\tau_L] \quad (11)$$

The first term in equation (11) reflects the fact that the bank can expect to receive her contribution to the default fund only in a good state of the economy with probability $\frac{1}{2}$. The second term expresses the return that is expected over the investment $(1 - n\alpha_{\bar{e}}^{CCP})$ minus dis-utility. The third term states the expected profit of the bank writing n identical contracts. The bank will always receive the transfer τ_H whenever the economy is in a good state but in case of a bad state of the economy, she will need to pay the transfer τ_L only when she has sufficient funds (i.e. with likelihood $\frac{1}{2}e$) in case her return on the asset is zero the transfer will be paid by the default fund.[17]

The corresponding incentive constraint (i.e. the condition to ensure that the bank prefers to exhibit high effort $e = a$) can be expressed as:

$$\frac{1}{2}n\tau(R_L) \leq (1 - n\alpha_{\bar{e}}^{CCP})(\theta_H - 1) \tag{12}$$

An agent proposes payment F where $\tau_R = F - R$ to maximize her utility:

$$U(R, F) \tag{13}$$

subject to

$$\pi(F, \alpha_{\bar{e}}^{CCP}, e) \geq \pi(0, 0, a) \tag{14}$$

Note that the agent does not need to consider the resource, the incentive constraint, or the number of contracts n the bank has or will sign since the CCP guarantees all contracts. The agent will however still need to consider the participation constraint and compensate the bank for opportunity costs of the default-contribution and probability of losing it.

Definition 3 *A CCP clearing equilibrium is an array $\langle n, \tau, \alpha_{\bar{e}}^{CCP}, e^* \rangle$ such that e^* maximize (11) given $\langle n, \tau, \alpha_{\bar{e}}^{CCP} \rangle$, τ maximize (13) given $\alpha_{\bar{e}}^{CCP}$ subject to (16), and $\bar{e} = e^*$.*

The solution to the agent's maximization problem is simple: she wants to receive the highest possible fixed payment F where the participation constraint is exactly met. This gives rise to the following results.

[17] The expected profit is based on the profit multiplied with the probability minus the disutility $(1 - n\alpha_{\bar{e}}^{CCP})C_e$:

$$\begin{array}{ll} \text{I: } e\frac{1}{2} & n\alpha_{e}^{CCP} + (1 - n\alpha_{\bar{e}}^{CCP})\theta_H - n\tau_H \geq 0 \\ \text{II: } (1-e)\frac{1}{2} & n\alpha_{\bar{e}}^{CCP} - n\tau_H \geq 0 \\ \text{III: } e\frac{1}{2} & (1 - n\alpha_{\bar{e}}^{CCP})\theta_H - n\tau_L \geq 0 \\ \text{IV: } (1-e)\frac{1}{2} & 0 \end{array}$$

Proposition 4 *A CCP clearing equilibrium exists with* $\bar{e} = a$ *in case of* $n \leq N_a^{CCP} = \lfloor 2\frac{1-(1-a)r_a}{\frac{1}{2}+(1-a)(\theta_H-1)}(\theta_H - 1)\rfloor$ *and* $\bar{e} = b$ *for* $N_a^{CCP} < n \leq N_b^{CCP} = \lfloor \frac{2\theta_H}{(1-r_a)} \rfloor$ *with the following swap rates:*

$$F_a^{CCP} = \frac{1}{2}(R_H + R_L) - \alpha_a^{CCP} r_a = \frac{1}{2 + 2(1-a)r_a} + R_L$$

$$F_b^{CCP} = \frac{1}{2}(R_H + R_L) - \frac{1}{2} r_a = \frac{1}{2}(1 - r_a) + R_L$$

Given that $F_a^{CCP} - F_b^{CCP} = \frac{r_a(a+(1-a)r_a)}{2(1+(1-a)r_a)} > 0$ it is clear that all agents are better off, when the CCP sets the limit $n \leq N_1^{CCP}$ to ensures that the bank exerts high effort $e = a$.

6.2 Termination clause: Maximization problem and equilibrium

In what follows, I will adjust the maximization problem described above for the case where a CCP uses a termination clause. To ensure that the bank exerts high effort $e = a$, the incentive constraint in equation (12) must be rewritten as follows:

$$(1-\lambda)\pi(n^*, F, \alpha_{\bar{a},T}^{CCP}, a) + \lambda\pi(0, 0, \alpha_{\bar{a},T}^{CCP}, a) \geq (1-\lambda)\pi(N, F, \alpha_{\bar{a},T}^{CCP}, b) + \lambda \quad (15)$$

On the left-hand side the bank keeps the number of contracts at the critical level n^*, i.e. exerts a high effort. In the first term the rating agency sends (correctly) a good signal and the bank can keeps the negotiated contract (see profit in equation (11)). In the second term, the rating agency (wrongly) sends a bad signal and the CCP cancels the contract of the bank, in which case $\pi(0, 0, \alpha_{\bar{a},T}^{CCP}, a) = n\alpha_{a,T}^{CCP} + (1 - n\alpha_{a,T}^{CCP})(\theta_H - C_a)$.

On the right-hand side the bank contracts with the maximum number N of counterparties and exerts low level of effort b. In the third term the rating agency falsely sends a good signal, so that the bank can make a profit of $\pi(N, F, \alpha_{\bar{a},T}^{CCP}, b) =$

$\frac{1}{2}(1 - N(F_{a,T}^{CCP} - R_H))$.[18] In the fourth and last term, the rating agency uncovers that the bank signed too many contracts and the CCP terminates all immediately. Finally, an agent proposes payment F to maximize her utility as shown in equation (13) subject to

$$(1-\lambda)\pi(n, F, \alpha_{\bar{a},T}^{CCP}, a) + \lambda\pi(0, 0, \alpha_{\bar{a},T}^{CCP}, a) \geq \pi(0, 0, a) \tag{16}$$

The solution to the agent's maximization problem is simple: she wants to receive the highest possible fixed payment F where the participation constraint is exactly met. This gives rise to the following results.

Proposition 5 *A CCP clearing equilibrium with termination clause exists with* $\bar{e} = a$ *in case of* $n \leq N_{a,T}^{CCP} = \lfloor \frac{2}{R_H - F_{a,T}^{CCP}} (\frac{r_a}{1-\lambda} + \frac{1}{2}) \rfloor$ *with the following swap rate:*

$$\begin{aligned} F_{a,T}^{CCP} &= \frac{1}{2 + 2(1-a)r_a} + R_L - \frac{1-\lambda}{\lambda}\alpha_{a,T}^{CCP} r_a \\ &= \frac{\lambda}{\lambda + r_a(1-\lambda)} \frac{1}{2 + 2(1-a)r_a} + R_L \end{aligned}$$

Given that $F_a^{CCP} - F_{a,T}^{CCP} = \frac{1-\lambda}{\lambda}\alpha_{a,T}^{CCP} r_a > 0$ it is clear that implementing the termination clause comes at a cost for the agent.

7 Bilateral and CCP Clearing

In this section I extend the analysis by giving the agents the choice to clear the contract bilaterally with a bank or via a CCP. I will consider situations where there is one or two CCPs available. Before I define the clearing equilibrium I summarize and simplify the results obtained so far.

[18] The expected profit is based on the profit multiplied with the probability. I can differentiate the following four cases where in the last two cases the profit is zero. Note too, that $b\theta_H = 1$.

$.5b$	$N\alpha_{a,T}^{CCP} + (1 - N\alpha_{a,T}^{CCP})\theta_H - N(F_{a,T}^{CCP} - R_H) \geq 0$
$.5(1-b)$	$N\alpha_{a,T}^{CCP} - N(F_{a,T}^{CCP} - R_H) \geq 0$
$.5b$	0
$.5(1-b)$	0

Table 1: Summary of incentive compatible swap rates and critical market size

	Swap Rate	**Critical Market Size**	**Approx.**
Bilateral Clearing			
no collateral	$F_1^* = \frac{1}{1+a} + R_L$	$N_1 = \lfloor 2(1+a)(\theta_H - 1) \rfloor$	$N_1 = 1$
term. clause	dito	$N_T = \lfloor \frac{2(1+a)}{a} \left(\frac{r_a}{(1-\lambda)} + \frac{1}{2} \right) \rfloor$	$N_T = 2$
full collateral	$F_2^* = \frac{1}{2(1+r_a)} + R_L$	No limit	-
CCP Clearing			
CCP, $e = a$	$F_a^{CCP} = \frac{1}{2+2(1-a)r_a} + R_L$	$N_a^{CCP} = \lfloor 2 \frac{1-(1-a)r_a}{\frac{1}{2}+(1-a)(\theta_H-1)} (\theta_H - 1) \rfloor$	$N_a^{CCP} = 1$
CCP, $e = a$ term. clause	$F_{a,T}^{CCP} = \frac{\lambda}{\lambda+r_a(1-\lambda)} \frac{1}{2+2(1-a)r_a} + R_L$	$N_{a,T}^{CCP} = \lfloor \frac{2}{R_H - F_{a,T}^{CCP}} (\frac{r_a}{1-\lambda} + \frac{1}{2}) \rfloor$	$N_{a,T}^{CCP} = 2$
CCP, $e = b$	$F_b^{CCP} = \frac{1}{2}(1 - r_a) + R_L$	$N_b^{CCP} = \lfloor \frac{2\theta_H}{(1-r_a)} \rfloor$	$N_b^{CCP} = 2$

The approximation is based on the assumption, that a is close to one, the return θ_H is in the range of $1.25 \geq \theta_H < 1.5$, and that $\lambda \geq 0.5$. If several values are possible, then the lowest is picked.

As can be derived from Table 1 the swap rates can be ranked in the following manner: $F_1^* > F_a^{CCP} > F_2^* > F_b^{CCP}$, i.e. a bilateral contract with no collateral pays the highest swap rate, followed by a CCP-cleared contract where the bank exerts high effort. A bilateral fully collateralized contract has a lower swap rate still. Finally, the swap rate cleared at a CCP where banks exert low effort is the lowest. The ranking of $F_{a,T}^{CCP}$ depends on the conditional probability λ.[19] In the following I will assume, that λ is sufficiently high but below one to ensure that

[19]For example, with $\lambda = \frac{1}{2}$ it is that $F_{a,T}^{CCP} < F_2^*$ and with $\lambda = 1$ it follows that $F_a^{CCP} = F_{a,T}^{CCP} > F_2^*$

$F_a^{CCP} > F_{a,T}^{CCP} > F_2^*$.
The ranking of swap-rates cannot be translated one-to-one into utility ranking in all cases. The following general statements can be made: $U(R, F_a^{CCP}) > U(R, F_{a,T}^{CCP}) > U(R, F_2^*) > U(R, F_b^{CCP})$ because in all four cases, the payment of the swap rate is guaranteed and the ranking follows directly from the size of the swap rate. The ranking of the utility of the contract with no collateral $U(R, F_1^*)$ depends on the risk-aversion of the agent c, the return of the risky asset θ_H, and the default rate $1 - a$ as will be shown in the discussion that follows.
The ranking of the critical market size is as follows: $N_1 \leq N_a^{CCP} \leq N_{a,T}^{CCP} \leq N_T \leq N_b^{CCP}$. The critical market size refers to the number of contracts a bank can sign at most without violating the incentive constraint (i.e. to ensure that $e = a$) or the resource constraint (i.e. in order to ensure that the bank can pay when her returns are high).
Finally, the approximation in the last column in table 1 is based on the assumption, that a is close to one, the return θ_H is in the range of $1.25 \geq \theta_H < 1.5$, and that $\lambda \geq 0.5$. More concretely, I have that $N_1 = N_a^{CCP} < N_{a,T}^{CCP} = N_T = N_b^{CCP}$. These are the values that I will use in the following discussion of clearing equilibrium.
When discussing an equilibrium, I will use the following definition.

Definition 4 *A clearing equilibrium for market size N where agents can clear bilaterally or through one CCP is an incentive compatible contract $\langle F, \alpha \rangle$ where no agent has an incentive to deviate and offer a different bilateral contract $\langle F, \alpha \rangle'$.*
In case of two CCPs the agent has additionally no incentive to deviate and clear at the other CCP.

Before starting the analysis I need to clarify what an incentive compatible contract $\langle F, \alpha \rangle$ is and how an agent can deviate. First, incentive compatible contract means that the bank has an incentive to accept the contract (participation constraint), that the bank exerts the expected level of effort $e \in a, b$ and that the expected transfers $\tau_R = F - R$ are subject to the resource constraint. In this sense table 1 sums up all such contracts given that the banks does not sign more contracts than specified by the critical market size. Therefore, when discussing an equilibrium I need to check whether the bank has an incentive and the possibility to sign more

contracts.
Second, in case of one CCP the agent can only deviate by proposing another *bilateral* clearing contract. To understand this, consider the following two situations. Given all agents clear with a CCP under a certain contract then an agent can only deviate by proposing to clear bilaterally with a contract of her choice. The agent cannot propose another CCP contract because the CCP sets the margins and therefore indirectly defines the contract. However, if all agents clear bilaterally, then an agent cannot deviate by proposing to clear through a CCP (even if a CCP were available in the background) because if only one agent clears with a CCP then there is no diversification advantage. In such a situation the agent can only deviate by proposing another bilateral contract. Therefore, when discussing an equilibrium I need to check whether an agent is better off by suggesting another incentive compatible bilateral clearing contract.
Finally, $\langle F_b^{CCP}, \alpha_b^{CCP} \rangle$ can never be an equilibrium because an agent can always deviate by proposing $\langle F_2^*, \alpha_2^* \rangle$ and be better off. All agents weakly prefer $\langle F_1^*, \alpha_1^* \rangle$ **with** a termination clause compared to the same contract without a termination clause. I will thus only consider in the analysis the following set of four contracts: $\langle F_1^*, \alpha_1^* \rangle$ (with termination clause), $\langle F_2^*, \alpha_2^* \rangle$, $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$, and $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$.
I will start with the analysis of the situation where the agent can choose between clearing a trade bilateral or through one CCP and then proceed to analyze the situation where two CCPs compete. All results are summarized in table (2).

7.1 Bilateral clearing and one CCP

For a market size $\boldsymbol{N > N_T = 2}$ the threat of contract termination does not deter the agent from signing too many contracts and so the contract $\langle F_1^*, \alpha_1^* \rangle$ cannot be part of the equilibrium, i.e. there are three equilibrium candidates that I need to consider: It is in case of bilateral clearing full collateralization $\langle F_2^*, \alpha_2^* \rangle$ as well as the two CCP Clearing contracts $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ and $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$.
Consider first $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ which would provide the highest utility of the remaining possible contracts (i.e. incentive or resource constraint compatible) for the agents. It is only then an equilibrium, when no agent has an incentive to deviate. Given that all other agents clear with the CCP, then an agent who prefers bilateral

clearing with a termination clause (and zero collateral) compared to CCP clearing ($U(R, F_a^{CCP}) < U(R, F_1^*)$) has an incentive to deviate. This strategy works because the bank that the bilaterally clearing agent signs a contract with will not accept a second contract cleared by a CCP because she does not want to risk receiving a downgrade by the rating agency. However, if all agents deviate and sign a bilateral contract with a termination clause then the bank has an incentive to sign too many contracts. Therefore $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ is only an equilibrium when $U(R, F_a^{CCP}) \geq U(R, F_1^*)$ which ensures that no agents wants to deviate.

The contract $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$ is an equilibrium even if agents would prefer the bilateral contract $\langle F_1^*, \alpha_1^* \rangle$. Given that all agent use a CCP with this particular contract then no agent has an incentive to deviate and offer a bilateral contract $\langle F_1^*, \alpha_1^* \rangle$ because the termination clause is ineffective in case of $N_{a,T}^{CCP} = N_T$. In other words, if a CCP allows the bank to sign $N_{a,T}^{CCP}$ contracts then deviating and offering a bilateral contract is not incentive compatible since the bank has an incentive to deviate and exert low effort.

Finally, the full collateralization $\langle F_2^*, \alpha_2^* \rangle$ is only a possible equilibrium when $U(R, F_2^*) \geq U(R, F_1^*)$, i.e. when agents are sufficiently risk averse or the return θ_H is not too high. The reason is, that if all agents propose $\langle F_2^*, \alpha_2^* \rangle$ then an agent prefers to deviate if $U(R, F_2^*) \leq U(R, F_1^*)$ and offer $\langle F_1^*, \alpha_1^* \rangle$ instead. This strategy is incentive compatible, because the bank still has an incentive to exert high effort a.

For market size $\boldsymbol{N \leq N_T = 2}$ two changes occur. First, the contract $\langle F_1^*, \alpha_1^* \rangle$ with termination clause is now an equilibrium in case where $U(R, F_2^*) < U(R, F_1^*)$. Second, the range for contract $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$ where it is an equilibrium narrows: it is only then a possible equilibrium when $U(R, F_{a,T}^{CCP}) \geq U(R, F_1^*)$. The reason is that the agent can deviate and offer a bilateral clearing contract $\langle F_1^*, \alpha_1^* \rangle$ if he is better off without having to fear that the bank will sign too many contracts and exert low effort b.

Finally, which of those four equilibria provide highest utility for the agents? In case of low risk aversion c, or high return on the risky asset θ_H the agent rather prefers to trade a contract bilaterally without collateral, instead of choosing a CCP, i.e. $U(R, F_a^{CCP}) < U(R, F_1^*)$. In this case the equilibrium with highest utility for the agent is $\langle F_1^*, \alpha_1^* \rangle$. If the risk aversion of the agent c is high or the return of the bank θ_H is not too high then the agents prefers to choose the CCP instead of the bilateral contracts, i.e. $U(R, F_a^{CCP}) > U(R, F_1^*)$. In such a case the equilibrium with highest utility for the agents is $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$. Thus $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$ and $\langle F_2^*, \alpha_2^* \rangle$ are possible but never the preferable equilibria.

Proposition 6 *When agents can choose to clear either bilaterally or with one CCP then $\langle F_b^{CCP}, \alpha_b^{CCP} \rangle$ cannot be an equilibrium. In case of $N \leq N_T$ all other bilateral or CCP equilibria are possible. When the market size is $N > N_T$ then the bilateral clearing contract without collateral $\langle F_1^*, \alpha_1^* \rangle$ is not an equilibrium (but the remaining three are). In all cases $e = a$.*

7.2 Bilateral clearing and two CCPs

I will now analyze the situation with two CCPs. First, note that if there are two CCPs offering the same contract than each can set a limit of $n = 1$ only. Setting a limit of $n = 2$ does not work because then in aggregate a bank could clear at both CCPs, exhaust their resources in case of a bad state of the economy R_L, so there would not be sufficient capital around to finance the transfers.[20]

Second, note that no CCP will choose to use the termination clause. This is because the termination clause is expensive and has positive externalities. To see this, suppose two CCPs share the market, both set the limit $n = 1$ and both use the termination clause. This would be sufficient to ensure that all bank exert high effort $e = a$ and choose to clear only at one CCP. However, one CCP has an incentive to deviate and not apply the termination clause but still keep the limit $n = 1$. In this case all agents would choose this CCP and the other CCP who applies the termination clause will never be chosen. This is equivalent to the case of having one CCP only in the market as discussed in the previous section. Therefore, the second CCP has an incentive not to apply the termination clause too. But then the banks will have an incentive to clear at both CCPs and thus exert low effort and so $\bar{e} = b$ with the swap rate F_b^{CCP}.

I start first with the market size being $N \leq N_1 = 1$. In this case, the solution is simple: agents can choose either to clear bilaterally or through a CCP. All

[20]Because the resource constraint in case of θ_H, R_L would be violated: $(1 - 4\alpha_a^{CCP})\theta_H - 4(F_a^{CCP} - R_L)) < 0$.

four equilibria $\langle F_1^*, \alpha_1^* \rangle$ (with termination clause), $\langle F_2^*, \alpha_2^* \rangle$, $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$, and $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$ are possible depending on the agent's risk aversion or size of risky return θ_H (see discussion in section 7.1).

In case of a market size $N_1 < N \leq N_T = 2$ it can be shown, that agents prefer to clear bilaterally using the termination clause and zero collateral or full collateralization (depending on the agent's risk aversion). CCP clearing cannot be an equilibrium because if all clear at the CCPs then as shown above, the CCPs will not apply a termination clause and end up in an equilibrium where $\bar{e} = b$ in which case an agent has always an incentive to deviate and bilaterally clear with full collateralization because $U(R, F_2^*) > U(R, F_b^{CCP})$.

If the market size grows to $N > N_T = 2$ then bilateral clearing with full collateralization $\langle F_2^*, \alpha_2^* \rangle$ is the two possible equilibrium.

Proposition 7 *When agents can choose to clear either bilaterally or with one of two CCPs then in case of $N \leq N_1$ agents will either clear bilaterally with the optimal contract $\langle F_1^*, \alpha_1^* \rangle$ and a termination clause or use the CCP to clear the optimal contract $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$. In both cases $e = a$.*
In case of $N_1 < N \leq N_T$ all agents will either prefer $\langle F_1^*, \alpha_1^* \rangle$ or $\langle F_2^*, \alpha_2^* \rangle$.
When the market size is $N > N_T$ and then there is only one equilibrium $\langle F_2^*, \alpha_2^* \rangle$.

Table 2: Possible Clearing Equilibria

<table>
<tr><th></th><th>$N \leq N_1$</th><th>$N \leq N_T = 2$</th><th>$N > N_T = 2$</th></tr>
<tr><td>One CCP</td><td colspan="2">$\langle F_1^*, \alpha_1^* \rangle$, $\langle F_2^*, \alpha_2^* \rangle$,
$\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$, $\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$</td><td>$\langle F_2^*, \alpha_2^* \rangle$,
$\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$,
$\langle F_{a,T}^{CCP}, \alpha_{a,T}^{CCP} \rangle$</td></tr>
<tr><td>Two CCPs</td><td>same as above</td><td>$\langle F_1^*, \alpha_1^* \rangle$,$\langle F_2^*, \alpha_2^* \rangle$</td><td>$\langle F_2^*, \alpha_2^* \rangle$</td></tr>
</table>

7.3 Implications

The main results of the discussion above can be summarized as follows. The contracts $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ or $\langle F_1^*, \alpha_a^* \rangle$ offer the highest utility for the agents. Which of those two carries the highest utility depends on the risk aversion and the size of the risky return θ_H (or alternatively the opportunity cost of collateral). Agents with low risk aversion facing banks with high returns on the risky asset rather prefer to clear bilaterally without collateral $\langle F_1^*, \alpha_1^* \rangle$. Agents with high risk aversion facing banks with low risky returns rather prefer to clear with the CCP $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ without a termination contract.

These two contracts, however, do not always form an equilibrium for two reasons. The first reason is competition between CCPs. In case of two CCPs offering clering for the same market the clearing contract $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ cannot be an equilibrium because banks can clear contracts at different CCPs and so circumvent the limits that CCPs might set.

The second reason is market size. For big markets $\langle F_1^*, \alpha_a^* \rangle$ does not work because banks would sign too many contracts unchecked and so $\langle F_a^{CCP}, \alpha_a^{CCP} \rangle$ remains as the only viable alternative in case there is no competition between CCPs. However, agents have an incentive to deviate, not use the CCP and clear bilaterally (without collateral) instead.

In a sense, CCPs have tried to solve this issue by effectively imposing mandatory clearing for certain liquid (and thus big) markets and allowing only one CCP to do the clearing, i.e. it is not possible to trade in a derivative market without using the dedicated CCP which in many cases is an integral part of the exchange market. For example, LCH launched in 1999 a CCP for interest rate swaps (IRS) with wide support from global banks, covering by the end of 2006 around 40% of the global inter-dealer IRS market (CPSS [2007]). However, this strategy might not work for contracts that are traded at many marketplaces globally where it is difficult to impose de-facto mandatory clearing or where competition between CCPs in different jurisdictions is allowing the banks to distribute their contracts. Another reason where self-organized mandatory clearing with one CCP is difficult to impose is in markets which have grown from small to a big in a relatively short time frame. In such cases the market participants might not react quickly enough and establish market practices or even set up a CCP taking into account the growing importance and market size. This has arguably happened in the credit-default rate market which has seen an almost tenfold increase between 2004 to 2007 where it roughly peaked at USD $67,2$ trillion at end-2007.

As described above, in such cases it can be argued that regulation of margin requirements in bilateral as well as CCP cleared markets can help to overcome the coordination problem. However, imposing mandatory clearing or collateral requirements in markets that are small reduces the utility of the market participants. In addition, CCPs would need to charge more collateral than is required to guarantee a default-free contract. From an financial stability point of view that might be an acceptable outcome because in practice it is difficult to establish which markets are small and which one are big. In addition, the size of the markets might change fast, making regulatory adjustments necessary over time.

8 Concluding remarks

In this paper, I study derivatives markets where risk-averse agents have the choice to trade contracts with risk-neutral banks through a central counterparty (CCP), where the level of collateral is set by the CCP, or clear them bilaterally where the level of collateral can be negotiated. An important feature of the model is that without collateral, the agents cannot control the size of the positions taken by the bank. As a result the bank might find it profitable to trade too many contracts, exert low levels of effort and subsequently default on her obligations.

I show that collateral discourages banks from trading too many contracts, encourages the bank to exert high levels of effort, and ensures that the obligations are always met. But collateral exhibits positive externalities and is costly, therefore even risk-averse agents would prefer not to use it. I show that contractual innovations (e.g. termination clauses) cannot solve the moral hazard issue and mimic the function of collateral satisfactorily. Regulation requiring agents to exchange minimum levels of collateral help to overcome this problem.

I show that a single CCP deals with the moral hazard problem by setting position limits. By pooling risks the CCP can offer a default-free contract at lower levels of collateral compared to bilateral clearing and thus ensure that agents prefer to use the CCP. However, position limits of a CCP do not work when she competes with other CCPs for the same market. Again, regulatory minimum collateral requirements help to overcome the moral hazard problem.

Appendix: Proofs

Proof of Corollary 1

To solve the agent's maximization problem given action $e = a$ by the bank I need to consider the participation and resource constraint in equations (3) and (4). Considering the (binding) participation constraint I can write:

$$\pi(0,0,a) - \pi(F,\alpha,a) = \alpha\Big(r_a + \frac{1}{2}(1-a)\Big) + \frac{1}{2}\Big((F - R_H) + a(F - R_L)\Big) = 0$$

For $\alpha = 0$ the above equation can be rewritten $F_1^* = \frac{1}{1+a} + R_L$. For $\alpha = F_2^* - R_L$ the solution is $F_2^* = \frac{1}{2(1+r_a)} + R_L$.

The resource constraint needs to be considered in case of θ_H, R_L only to ensure that the bank can make a transfer when her return is high, which boils down to $(F_1^* - R_L) \leq \theta_H$ or just $\frac{1}{1+a} \leq \theta_H$, which is always satisfied given the range of values I assume for a and θ_H.

Finally, the agent will choose zero collateral in when her utility is higher than with full collateral, i.e.

$$\frac{1}{2}(1+a)u(F_1^*) + \frac{1}{2}(1-a)u(R_L) \geq u(F_2^*)$$

If the above inequality is not satisfied then the agent will choose full collateral.

Proof of Corollary 2

I will show that the agent is always better off with an optimal contract $\langle F_1^*, \alpha_1^* = 0\rangle$ than no contract at all, i.e. $U(R, F_1^*) > U(R)$ for all $a \in (0,1]$. The inequality can be explicitly expressed as follows.

$$-\frac{1}{2}(1+a)e^{-cF_1^*} - \frac{1}{2}(1-a)e^{-cR_L} \geq -\frac{1}{2}e^{-cR_H} - \frac{1}{2}e^{-cR_L}$$

After multiplying the equation with e^{cR_L} and some further reformulation I get:

$$-(1+a)e^{-\frac{c}{1+a}} + a > -e^{-c}$$

Now it can be shown, that in case of $a = 0$ the agent is equally well of with or without the contract i.e. $-e^{-c} = -e^{-c}$ and that taking the derivative w.r.t. c of the difference

$-(1+a)e^{-\frac{c}{1+a}} + a + e^{-c}$ equals

$$\frac{\partial \cdot}{\partial a} = 1 - e^{-\frac{c}{1+a}}(c + \frac{c}{1+a}) > 0$$
$$\Leftrightarrow ln(c + \frac{c}{1+a}) < \frac{c}{1+a}$$

Which proofs that increasing a always increases the utility of the agent with the contract and so the inequality is valid for any $a > 0$.

Proof of Proposition 4

Given that each bank signs the maximum allowed $n = n_{CCP}$ contracts, all agents will propose to a bank a contract they will accept. The (binding) participation constraint from equation (16) for any $e \in \{a, b\}$ can be written as:

$$\underbrace{n\alpha_e^{CCP} - n(1-p)\alpha_e^{CCP}}_{=np\alpha_e^{CCP}} + (1 - n\alpha_e^{CCP})\underbrace{(1 + r_e)}_{=\theta_H e - C_e} - n[\frac{1}{2}\underbrace{(F_e^{CCP} - R_H)}_{\tau_H} + \frac{1}{2}e\underbrace{(F_e^{CCP} - R_L)}_{\tau_L}] = \underbrace{1 + r_a}_{\theta_H a - C_a}$$

After some reformulations I get $F_e^{CCP} = \frac{1}{2}(R_H + R_L) - \alpha_e^{CCP} r_e + \frac{1}{n}(r_e - r_a)$ so that:

$$F_a^{CCP} = \frac{1}{2}(R_H + R_L) - \alpha_a^{CCP} r_a$$
$$F_b^{CCP} = \frac{1}{2}(R_H + R_L) - \frac{1}{2} r_a$$

Note that both n as well as α_e^{CCP} are exogenous.

Proof of Proposition 5

The (binding) participation constraint from equation (16) can be written as:

$$\lambda\Big[\frac{1}{2}n\alpha_{a,T}^{CCP} + (1 - n\alpha_{a,T}^{CCP})(1 + r_a) - n[\frac{1}{2}(F_{a,T}^{CCP} - R_H) + \frac{1}{2}a(F_{a,T}^{CCP} - R_L)]\Big]$$
$$+(1-\lambda)\Big[n\alpha_{a,T}^{CCP} - (1 - n\alpha_{a,T}^{CCP})(1 + r_a)\Big] = 1 + r_a$$

After some reformulations I get $F_a = \frac{1}{2}(R_H + R_L) - \alpha_{a,T}^{CCP} r_a - \frac{1-\lambda}{\lambda}\alpha_{a,T}^{CCP} r_a$

Proof of Proposition 6

I want to show that for certain parameters $U(R, F_a^{CCP}) \geq U(R, F_1^*, \alpha_1^* = 0)$, even if termination clause is sufficient to ensure that $e = a$ (i.e. when $N \leq N_T$). To do so, I will use the utility $u(x) = -e^{-cx}$. Each expected utility can be expressed as

$$U(R, F_1^*) = (1 - \frac{1}{2}(1-a))u(F_1^*) + \frac{1}{2}(1-a)u(R_L)$$
$$U(R, F_a^{CCP}) = u(F_a^{CCP})$$

So that the above inequality can be written as.

$$-e^{-\frac{c}{2+2(1-a)r_a}} \geq -(1 - \frac{1}{2}(1-a))e^{-\frac{c}{1+a}} - \frac{1}{2}(1-a)$$
$$e^{-\frac{c}{2+2(1-a)r_a}+\frac{c}{1+a}} \leq 1 + \frac{1}{2}(1-a)(e^{\frac{c}{1+a}} - 1)$$
$$\frac{c(1+2r_a)}{(1+a)(1+(1-a)r_a)} \leq e^{\frac{c}{1+a}} - 1$$

A decrease in θ_H (which reduces r_a) or an increase of risk aversion c makes CCP clearing more attractive.

Second, I want to show that for $N > N_T$, i.e. when in a bilateral market full collateralization is necessary to ensure $e = a$, the agent always prefers CCP clearing. In other words, I want to show that $U(R, F_a^{CCP}) > U(R, F_2^*)$, where $F_2^* = \frac{1}{2(1+r_a)} + R_L$, so that $F_2^* - F_a^{CCP} = \frac{1}{2}(\frac{1}{1+r_a} - \frac{1}{1+r_a-ar_a}) < 0$.

Thus, if all other agents clear with a CCP with clearing limit $n = 1$, then an agent who prefers bilateral clearing will propose a

Proof of Proposition 7

Agent prefers bilateral clearing if full collateralization necessary compared to two CCPs and ($\bar{e} = b$)

$F_{CCP}^b = \frac{1}{2}\Delta R - \frac{1}{2}r_a + R_L$, and in case of full collateralization $F_a = \frac{1}{2(1+r_a)}\Delta R + R_L$, so that

$$F_a - F_{CCP}^b = \frac{1}{2}r_a - \Delta R\frac{r_a}{2(1+r_a)} = \frac{r_a^2}{2(1+r_a)} > 0$$

Which means, that full collateralization in a bilateral market is better than low effort in CCP clearing.

Agents prefer bilateral clearing if $e = a$ but $\bar{e} = b$ (proof)

I want to show that $U(R, F^b_{CCP}) \leq U(R, F_a, \alpha^*_a = 0)$, where $F^b_{CCP} = \frac{1}{2}\Delta R - \frac{1}{2}r_a + R_L$ since $n = 2$ (important! Here I assume that the bank clears at both CCPs) and where $F_a - F^b_{CCP} = \frac{1-a}{2(1+a)}\Delta R + \frac{1}{2}r_a$:

$$
\begin{aligned}
-e^{-F^b_{CCP}} &\leq -(1 + r_a)e^{-F_a} \\
e^{F_a - F^b_{CCP}} &\geq (1 + r_a) \\
\frac{1-a}{2(1+a)}\Delta R + \frac{1}{2}r_a &\geq r_a \\
\frac{1-a}{(1+a)}\Delta R &\geq r_a
\end{aligned}
$$

which can be satisfied.

References

R. Angelo, P. Schaffner, and M. Vasios. Regulatory effects on short-term interest rates. *Bank of England Staff Working Paper*, 801, 2019.

N. Arora, P. Gandhi, and F. A. Longstaff. Counterparty credit risk and the credit default swap market. *Journal of Financial Economics*, 103(2):280 – 293, 2012. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2011.10.001. URL `http://www.sciencedirect.com/science/article/pii/S0304405X11002327`.

B. Biais, F. Heider, and M. Hoerova. Risk-sharing or risk-taking? counterparty risk, incentives, and margins. *The Journal of Finance*, 71(4): 1669–1698, 2016. ISSN 1540-6261. doi: 10.1111/jofi.12396. URL `http://dx.doi.org/10.1111/jofi.12396`.

V. Bignon and G. Vuillemey. The failure of a clearinghouse:empirical evidence. Working papers 638, Banque de France, 2017. URL `https://ideas.repec.org/p/bfr/banfra/638.html`.

A. Bisin and D. Guaitoli. Moral hazard and nonexclusive contracts. *The RAND Journal of Economics*, 35(2):306–328, 2004. ISSN 07416261. URL `http://www.jstor.org/stable/1593693`.

D. S. Bizer and P. M. DeMarzo. Sequential banking. *Journal of Political Economy*, 100(1):41–61, 1992. ISSN 00223808, 1537534X. URL `http://www.jstor.org/stable/2138805`.

A. R. Cenedese Gino and M. Vasios. Otc premia. *Bank of England Staff Working Paper*, (751), Aug. 2018.

R. Cont and T. Kokholm. Central clearing of otc derivatives: Bilateral vs multilateral netting. *Statistics & Risk Modeling*, 31:3–22, 2014. doi: https://doi.org/10.1515/strm-2013-1161.

CPSS. New developments in clearing and settlement arrangements for otc derivatives. Technical report, Bank for International Settlements, Mar. 2007.

D. Duffie and H. Zhu. Does a central clearing counterparty reduce counterparty risk? *The Review of Asset Pricing Studies*, 1(1):74–95, 2011. doi: 10.1093/rapstu/rar001. URL + `http://dx.doi.org/10.1093/rapstu/rar001`.

FSB-BIS-CPMI-IOSCO. Incentives to centrally clear over-the-counter (otc) derivatives. Technical report, Financial Stability Board, Aug. 2018.

W. Huang. Central counterparty capitalization and misaligned incentives. BIS Working Papers 767, Bank for International Settlements, Feb. 2019. URL `https://ideas.repec.org/p/bis/biswps/767.html`.

Y. Leitner. Inducing agents to report hidden trades: A theory of an intermediary*. *Review of Finance*, 16(4):1013–1042, 2012. doi: 10.1093/rof/rfr017. URL + `http://dx.doi.org/10.1093/rof/rfr017`.

O. Lewandowska. Otc clearing arrangements for bank systemic risk regulation: A simulation approach. *Journal of Money, Credit and Banking*, 47 (6):1177–1203, 2015. ISSN 1538-4616. doi: 10.1111/jmcb.12241. URL `http://dx.doi.org/10.1111/jmcb.12241`.

M. Maizar, M. Jaquet, and T. Spillmann. Credit and counterparty risk: Why trade under an isda with a csa? *Schweizerische Zeitschrift fuer Gesellschafts- und Kapitalmarktrecht sowie Umstrukturierungen*, (1):63–71, 2010.

T. Santos and J. A. Scheinkman. Competition among Exchanges. *The Quarterly Journal of Economics*, 116(3):1027–1061, 08 2001a. ISSN 0033-5533. doi: 10.1162/00335530152466304. URL `https://doi.org/10.1162/00335530152466304`.

T. Santos and J. A. Scheinkman. Financial intermediation without exclusivity. *American Economic Review*, 91 (2):436–439, May 2001b. doi: 10.1257/aer.91.2.436. URL `https://www.aeaweb.org/articles?id=10.1257/aer.91.2.436`.

G. Vuillemey. The value of central clearing. *The Journal of Finance*, n/a(n/a), 2020. doi: 10.1111/jofi.12902. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12902`.

Selbständigkeitserklärung
Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 14. Juli 2020

Robert Oleschak